DEVELOPMENT AND EVALUATION OF THE DIGIT TRIPLET AND AUDITORY-VISUAL MATRIX SENTENCE TESTS IN MALAY

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy in Audiology

in the University of Canterbury

by Saiful Adli Jamaluddin

University of Canterbury

2016

TABLE OF CONTENTS

ABSTRACT

The objective of this study is to develop and evaluate versions of the digit triplet (MDTT) and matrix sentence tests in Malay (MMST-AV). The development of both tests involved the selection, recording and normalisation with level adjustments of speech stimuli in the form of digits and words. The MDTT was developed for headphone and telephone applications in test specific noise (TSN) and spectrotemporal gap noise (STG). The MMST-AV was developed for headphone application in TSN and 6-talker babble noise (BN). To allow for auditory-visual (AV) mode of assessment of the MMST-AV, a visual component was added which required additional studies to investigate the optimal method of normalisation and the refinement of video samples. Both tests were evaluated auditorily for list equivalency in normal hearing listeners and were further validated in a group of listeners with varying hearing levels. Additionally, the evaluation of the MMST-AV included the investigation of the training effects. Eight lists of 27 digit triplets and 15 lists of 30 sentences were evaluated monaurally in a closed-set response format for the MDTT and MMST-AV, respectively. A total of 166 normal hearing and 26 hearing impaired participants were recruited for this study. For the MDTT, evaluation in fixed SNRs resulted in a mean speech reception threshold (SRT) of -11.3 ± 0.34 dB SNR for headphone application in TSN; -11.9 ± 0.4 dB SNR for headphone application in STG; -10.24 ± 0.1 dB SNR for telephone application in TSN and -10.8 ± 0.3 dB SNR for telephone application in STG. The mean SRT and slope normative reference for the MMST-AV were -10.1 ± 0.2 dB SNR and 14.9 ± 1.2 %/dB, respectively in TSN whereas in BN, mean SRT and slope were –6.4 ± 0.2 dB SNR and 12.2 ± 0.7 %/dB, respectively. A significant training effect of 1.4 dB was observed for the first two consecutive measurements in the TSN and 0.8 dB in BN. Evaluation in listeners with varying hearing levels in the MDTT revealed test sensitivities and specificities of more than 85% in all four test conditions. Performances of normal and hearing impaired groups were found to be equal in the AV and visual-only mode of testing after the effect of participant’s age was controlled. In conclusion, both the MDTT and MMST-AV showed good agreement between the SRTs and slope to other versions of the tests and are suitable for repeated measurements.

ACKNOWLEDGEMENTS

To all readers, May peace be upon you or *"Salam sejahtera"*in Malay.

I would like to acknowledge that the production of this dissertation would be not possible without the help of so many great people around me for whom I am truly grateful.

My most sincere appreciation to my senior supervisor, Associate Professor Greg O'Beirne, for his continuous support and advice. I am fortunate to have been guided by such a brilliant person. Thank you for your advice and enduring patience. This experience will certainly have a lasting impression on me.

Thank you to the Ministry of Higher Education Malaysia, International Islamic University Malaysia and GN Resound for providing financial support for this study.

My humble gratitude to the many participants for their time and effort to take part in the long and strenuous tests.

To Dr. Don Sinex for his input especially during the preparation of my proposal.

To Dr. Don Clucas & Mr. John Chrisstoffels for lending me your expertise.

I am truly grateful for the unwavering support of my wife, Nurlin Ali Hanafiah who's been my emotional and intellectual partner throughout this journey.

My heartfelt appreciation to my parents and parents-in-law for their support and prayer and particularly for looking after the kids when in need.

A special mention to Sarah Rahmat, Dr. Mazidah Ahmad Mansur and Dr. Alison M .Johnston for being my friends and mentors throughout this study.

Last but not least, thank you to the staff and students of the Department of Audiology and Speech-Language Pathology, International Islamic University Malaysia for their support during periods of data collection.

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

A	Auditory-only
ABR	Auditory brainstem response
AE	Auditory enhancement
AI	Articulation index
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
ASHA	American Speech–Language–Hearing Association
AV	Auditory-visual mode
BM	Basilar membrane
CAPD	Central auditory processing disorder
CF	Characteristic frequency
dB HL	Decibels hearing level
dB SPL	Decibel sound pressure level
DRT	Direct realist theory
DTT	Digit triplet test
EC	Earcheck screening test
ERB	Equivalent rectangular bandwidth
ERP	Evoked related potentials
FA	Freidreich's Ataxia
FLMP	Fuzzy logic model of processing
***f*MRI**	Functional magnetic resonance imaging
GA	General approach
GPS	Global positioning system
HHIE	Hearing handicap inventory for elderly
HINT	Hearing in noise test
HTL	Hearing threshold level
ICRA	International Collegium of Rehabilitative Audiology
IE	Integration enhancement
IHC	Inner hair cells
IIUM	International Islamic University Malaysia

LTASS	Long term averaged speech spectrum
MDTT	Malay digit triplet test
MEM	Middle ear muscle reflex
MLP	Maximum likelihood procedure
MMN	Mismatch negativity
MMST-AV	Malay auditory-visual matrix sentence test
MOC	Medial olivocochlear complex
MST	Matrix sentence test
MT	Motor theory of speech perception
NHT	National hearing screening test
NZD	New Zealand Dollar
NZHST	New Zealand Hearing Screening Test
OHC	Outer hair cells
OLSA	Oldenburger Satztest/Oldenburg sentence test
PC	Personal computer
PCM	Pulse code modulated
PTA	Pure tone audiometry
QALY	Quality adjusted-life-year
RM	Malaysian Ringgit
RM-ANOVA	Repeated measures analysis of variance
RMS	Root mean square
ROC	Receiver-operator curve
SAC	Self-assessment of communication
SDT	Speech detection threshold
SNR	Signal-to-noise ratio
SRT	Speech reception threshold
SRTn	Speech reception threshold in speech-in-noise test
STG	Spectrotemporal gap noise
TEN	Threshold equalizing noise test
TSN	Test specific noise
UCAST	University of Canterbury Adaptive Speech test
UC-AVMST	University of Canterbury Auditory-Visual Matrix Sentence Test

V	Visual-only
VE	Visual enhancement
VI	Virtual interface
VOT	Voice onset time
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background

One of the significant difficulties faced by hearing impaired listeners is listening in the presence of background noise. Reduced frequency and temporal resolution abilities in listeners with sensorineural hearing loss contribute to the distortion and deterioration of the acoustic signal, such that a decrease in signal-to-noise ratio affects overall intelligibility to a greater extent than in normal hearing listeners (Moore, 1996). Therefore, in order to assess an individual's real-world speech perception ability it is appropriate to utilize speech materials such as sentences in noise instead of pure tone signals or monosyllabic words in quiet. There are two reasons for this: Firstly, speech signals are more representative of real-world communication and secondly, the use of noise increases the sensitivity of the test in identifying significant hearing impairments.

The World Health Organization estimates that there are 360 million persons (5.3% of world population) with disabling hearing loss in the world and it is estimated that 91% of them are adults. The prevalence increases with age where one out of three adults over the age of 65 has disabling hearing impairment (WHO, 2012). Therefore, hearing loss should be viewed as a major world health issue that needs to be addressed in terms of early identification, intervention and rehabilitation. The recent development of two particular tests has allowed audiological screening and diagnostic tests to be standardized, which is a remarkable accomplishment considering the multiple languages that the test has been produced in in recent years. These tests are known as the "digit triplet" and "matrix sentence" tests. Both tests use limited collections of speech items to form a full test from either three digit pairs called triplets or complete sentences with a fixed sentence structure. This enables the test to be both homogeneous in intelligibility and comparable across languages (Akeroyd et al., 2015). The digit triplet test was initially introduced in Dutch by Smits and colleagues for delivery via telephone (Smits, Kapteyn, & Houtgast, 2004), and they reported high sensitivity and specificity in the detection of sensorineural hearing loss. In a later study (Smits, Kramer, & Houtgast, 2006), they also reported that the digit triplet test was a more accurate and reliable screening tool than a short disability-type questionnaire as it is not biased by age or the misperception of hearing disability. Versions of the digit triplet test have been adapted to many languages and also delivered via the internet using headphones (monaural testing) or loudspeakers (binaural testing).

The matrix sentence test was first introduced by Hagerman (1982) in Swedish where ten sentences with the same syntax were edited to form new unique sentences with the same structure. The biggest advantage of this type of test is that similar reproduction can be applied to other languages. Zokoll and colleagues discussed how these tests are highly comparable with each other due to three major factors by examining the versions of matrix sentence tests in Russian, Spanish, Swedish, Polish, Turkish, British English, French and Danish (Zokoll et al., 2013). The first is that the tests were designed to retain the same language complexity and structure; secondly, the tests used adaptive measures to track responses of participants which enables it to converge to the level at which 50% word intelligibility is achieved. Finally, the matrix sentence tests were methodically designed using the same process which includes the process of optimizing the speech materials to ensure equal intelligibility. Acknowledging the emerging trend in developing this flexible tool, the International Collegium of Rehabilitative Audiology (ICRA) have produced a review to recommend measures to be taken to design, evaluate and validate the digit triplet and the matrix sentence tests so that other attempts to develop this test in other languages can be compared internationally (Akeroyd et al., 2015).

Currently, there are very few published and validated speech tests that are available in the Malay language. This study proposes the development of two speech-in-noise tests in the Malay language for the adult population: the digit triplet test which is to be used for hearing screening purposes, and a matrix sentence test to be used for diagnostic purposes. The digit triplet test offers a unique opportunity for the adult Malay speaking population to self-administer a hearing screening test and objectively evaluate their hearing, as well as motivate them to seek hearing consultation with hearing professionals. The matrix sentence test on the other hand has the potential to be used as a flexible diagnostic tool for audiologists due to its unpredictable fixed sentences and ability to change type of response methods and presentation modes. The scope of this thesis will include the development of test materials, as well as normalisation and evaluation of both tests. The outcome of this study is expected to assist the Malaysian public and hearing care professionals in Malaysia in providing a more comprehensive audiological service.

In this chapter, keywords and common references that are frequently used in studies related to speech perception testing in noise are described. In addition, the main purpose of study is explained to help readers understand the current need for both tests in the Malaysian context. It is hoped that with the explanation provided, the significance of this study is adequately justified. The general and specific aims of this study are also described for both tests.

In Chapter 2, reviews of current and relevant literature are discussed. Readers are first introduced to fundamental elements in speech and its perception including auditory-based theories as well as auditory-visual theories of speech perception. The middle part of this chapter aims to introduce and discuss different speech perception tests and the various elements that could affect the results obtained. A substantial part of the review centres on the development and implementation of digit triplet and matrix sentence tests in various languages as well as speech testing in the Malaysian context.

As the development of test construction for both digit triplet test (DTT) and matrix sentence test (MST) are similar, it is jointly described in Chapter 3. This includes the topics on background behind speech material selection, as well as recording and editing of the test materials. Also described in this chapter are methods relating to the construct of the background noise which also include the design and development of the spectral and temporal gap noise for the Malay DTT (MDTT). Readers are also introduced to a different approach to editing speech materials for the matrix sentence test where visual stimulation is added to the test thereby allowing an auditory-visual mode of testing.

Chapter 4 describes the normalisation or optimization processes for the Malay DTT. Normalisation was done using headphone and telephone in two different background noises which are the test specific noise and the spectrotemporal gap noise. Lists of equally intelligible triplets were generated and were used for the evaluation process.

Chapter 5 describes and discusses various aspects of the normalisation process of the Malay auditory-visual matrix sentence test (MMST-AV) which includes level adjustments of speech stimuli, evaluating scoring and normalisation methods introduced in the New Zealand English auditory-visual matrix sentence test as well as investigating the quality of the visual recording of the test. This is due to differences in recording techniques which have been changed to accommodate provision of the auditory-visual testing mode.

Evaluation of lists to examine their equivalency is described for both the Malay DTT and Malay MST in Chapter 6 using both adaptive and fixed signal-to-noise ratio (SNR) measurements. Additionally, learning effects observed in the Malay MST are also discussed in this chapter.

Both tests were evaluated in terms of their specific configuration of testing and are described in Chapter 7. Listeners with varying hearing levels were evaluated adaptively to compare between speech reception thresholds in noise and pure tone audiometric thresholds. Correlations between the Malay DTT and MST are included in this chapter. To investigate the benefit of visual input to listeners with normal and impaired hearing in difficult listening conditions using the auditory-visual function of the Malay matrix sentence test, forty six listeners were tested in auditory alone, visual alone and auditory-visual modes. Comparisons between different modes are discussed in this chapter.

Concluding remarks which cover all experiments within the scope of this study are described in Chapter 8. Also noted in this chapter are limitations of this study as well as recommendations for future research.

1.2 List of definitions

Speech	A vocalized form of communication that is presented via a language which is governed by certain rules.
Communication	An exchange of information and ideas, needs and desires between speaker and listener.
Disorder	An abnormality in human anatomy and/or physiology.
Disability	Any restriction or lack of ability to perform an activity.
Hearing loss	Being partly or totally unable to hear sound in one or both ears.
Malay language	An Austronesian language spoken in Malaysia, Indonesia, Singapore, Brunei and Thailand.
Matrix sentence test (MST)	A speech-in-noise test that is presented in fixed syntactical but semantically unpredictable sentences formed using fifty predetermined words. This test is typically presented to adults and used for hearing diagnostics and rehabilitation.
Digit triplet test (DTT)	A speech-in-noise test that is presented as randomized three digits pairs or triplets. This test is commonly used for hearing screening purposes.
Signal-to-noise ratio (SNR)	A measure that compares the level of a desired signal to the level of background noise. It is defined as the ratio of signal power to the noise power, often expressed in decibels.
Speech reception threshold in noise (SRTn)	Speech-to-noise ratio in decibels required for fifty percent of the speech stimuli to be identified correctly.
Slope	The percentage of score over the difference in decibels at the level of the SRTn

1.3 Problem statement

Evidence suggests that many individuals with hearing impairment would only seek help if they perceive their hearing to be poor (Meyer & Hickson, 2012). Rosdina et al. (2010) studied 114 elderly adults in Malaysia and discovered a significant correlation between self-reported hearing loss and pure tone four frequency average hearing thresholds (250, 500, 200 & 4000 Hz) of more than 40 dB HL. This may suggest that by the time a person acknowledges a hearing loss, the level of impairment is already disabling (WHO, 2012) and could possibly limit the success of a rehabilitation programme. In Malaysia, standardized versions of the Hearing Handicap Inventory for the Elderly (HHIE) or the Self Assessment of Communication (SAC) are not available and a visit to a hearing care centre is necessary to obtain information regarding a person's hearing, which is not always accessible to some people. Traditional hearing screening methods are labour and equipment intensive. They require a trained individual to conduct the test as well as adequate and calibrated audiometric equipment (ASHA, 1996). A low level of ambient noise is also important in order to obtain an accurate and reliable outcome. Therefore, there is a need for an objective hearing test for screening purposes that is self-administered and accessible to all Malay speakers. A test using speech materials such as digits could offer an easy way for an individual to perform such a test. A Malay digit triplet test is applicable to most Malaysians (even those who are non-native Malay speakers) because digits are used commonly in everyday speech, offer no contextual clues, and can be adapted into automatic telephone or internet-based tests.

For many years, some audiologists in Malaysia have had to resort to using materials in English to compensate for the limited availability of speech test materials in Malay, but non-native speakers may not perform as well as native speakers of the language (Warzybok, Brand, Wagener, & Kollmeier, 2015). Many choose to leave out this test altogether, either for convenience or because of the lack of necessary equipment. At the moment, speech testing in Malay is limited to central auditory processing disorder (CAPD) investigation (Mukari, Keith, Tharpe, & Johnson, 2006) and hearing aid and cochlear implant verification through the Malay Hearing-in-Noise Test (HINT) (Quar et al., 2008). A valid speech-in-noise test should be made accessible to clinics throughout Malaysia to improve the reliability and broaden the scope of audiological intervention and rehabilitation. An everyday sentence speech test such as the HINT (Nilsson, Soli, & Sullivan, 1994) is semantically predictable and is susceptible to patients memorizing the entire sentence, and up to a year is needed before the test can be used again (Hochmuth et al., 2012). This particular issue will not affect the matrix sentence test as previously developed matrix sentence tests have shown high accuracy and low redundancy due to its semantically unpredictable structure which enables it to be administered to the same patient several times without the concerns of learning effects (Wagener & Brand, 2005). In addition, matrix sentence tests can be designed to be a flexible tool where the signal can be presented with or without noise and the test administered in an open or closed response method. Therefore, it is proposed that a Malay matrix sentence test would cater for the need for a diagnostic speech in noise test for the adult population.

1.4 Significance of research

In Malaysia, access to an audiologist and complete audiological tools are limited as there were only 295 registered audiologists in the Malaysian National Society of Audiologists in 2015 (Farah Wirda, personal communication, January 22, 2016). Additionally, presumably due to poor hearing awareness and the relatively high cost of hearing aids, it is possible that these factors could jointly hinder people with hearing impairment from seeking help (Carson, 2005). The development of the Malay digit triplet test (MDTT) is expected to assist in the provision of valid speech intelligibility tests as an accessible and inexpensive tool for hearing screening in Malaysia, as well as create better hearing awareness.

Due to the limited number of available speech in noise tests in Malaysia, the Malay matrix sentence test could provide a flexible tool for audiologists to conduct diagnostic or rehabilitation services for the adult population. Future development of the materials can be applied to the paediatric population and other specific tests of hearing disorders such as the assessment of auditory neuropathy spectrum and central auditory processing disorders.

1.5 Research aims & objectives

The main aim this study was to develop and validate two speech-in-noise tests in Malay, specifically;

Test 1: Digit triplet test via internet and telephone for hearing screening applications.

Test 2: Auditory-visual matrix sentence test for diagnostic hearing assessment.

1.5.1 The specific objectives for the MDTT development were to:

- Record and edit speech recordings to produce digit triplets;
- Generate two types of masking noise using the edited digit triplets, which are:
    - Steady state speech shaped noise, also known as test specific noise (TSN); and
    - Spectral and temporal gap (STG) noise based on the TSN.
- Normalise the MDTT for both headphone and telephone applications;
- Produce and evaluate unique and equivalent digit triplet lists for both types of transducers and noises using fixed and adaptive SNR measurement;
- Generate and analyse receiver-operator curves (ROCs) to identify specific cut-off regions for pass and refer criterions by testing normal and hearing impaired participants using a multi-centre testing approach;
- Discuss the effects of transducer and noise type in the MDTT; and
- Identify the sensitivity and specificity of the Malay digit triplet test for the detection of sensorineural hearing loss in both types of application.

1.5.2 The specific objectives for the MMST-AV were to:

- Record, edit and enhance audio and video recordings of a selected speaker to produce audio-visual materials for the MMST-AV;
- Generate two types of masking noise, which are:
    - Steady state speech shaped noise; and
    - 6-talker babble noise.
- Normalise the MMST-AV audio recordings;
- Identify the appropriate method of normalisation;
- Identify naturalness of video edits and exclude visually non-natural sentences;
- Produce and evaluate unique and equivalent lists for the test in both types of noise;
- Identify learning effects in the MMST-AV; and
- Measure and discuss the performances of participants with varying levels of hearing in the auditory only, visual only and auditory-visual modes of the MMST-AV using a multi-centre testing approach.

1.6 Hypotheses

The hypotheses are as follows:

1.6.1 There are no significant differences in SRTn between the lists for both MDTT and MMST-AV.

1.6.2 There are significant and positive correlations between MDTT and MMST-AV results and participants' hearing thresholds.

1.6.3 The MDTT and MMST-AV have high sensitivity and specificity in identifying sensorineural hearing loss.

1.6.4 Using temporally modulated noise (STG noise) allows for release from masking for normal hearing participants leading to improved SRTn in the MDTT.

1.6.5 Using 6-talker babble is detrimental to SRTn scores compared to test specific noise in the MMST-AV.

1.6.6 The SRTn and slope scores obtained from normal hearing listeners are comparable to other established matrix sentence tests.

1.6.7 There are significant differences between results obtained from the normal and hearing impaired groups using different types of masker.

1.6.8 Both the normal and impaired hearing groups show varied responses in the visual alone and the auditory-visual tests.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to speech perception

Speech perception can be defined as the ability to collect, organize, identify and interpret human verbal information into meaningful signals. Speech perception is a complex process where information is managed from a combination of acoustic waveforms, linguistic and contextual cues and visual information. Due to its complexity, the auditory processes are not fully understood despite much research into speech perception (Moore, 2003). Because speech is such an integral part of daily communication, the evaluation of speech perception is necessary to gauge a listener's ability to identify, discriminate, comprehend and understand speech, especially in the hearing impaired population. This section of the literature review attempts to briefly discuss the complex nature of speech perception abilities in both normal and hearing impaired listeners. This includes the physical attributes of speech, the physiology of speech perception and factors that affect it.

2.1.1 The physics of speech

To better understand how speech is perceived, it is necessary that we attempt to understand how speech is produced and the physical attributes that it yields. In this part of the review, a brief explanation on the characteristics of speech sounds will be made. Speech sounds can be broadly classified as vowels and consonants and in its most basic form, speech sounds are referred to phonemes, which is the smallest meaningful unit of a language's sound systems. For example, the English words *ball* and *call* are differentiated by the phonemes /b/ and /k/. The exchange of these two phonemes results in the change of the sound of the word hence altering its meaning. Acoustically, each phoneme can be generally characterized by its intensity, duration and overall spectral content. The spectral and time analysis of speech sounds can be observed using a spectrogram (Figure 1). The vertical axis represents the frequency domain and the horizontal axis is the time domain. Another dimension that is measured using the spectrograph is the intensity of speech sound which is represented by the shade of colour within the frequency-time domain.

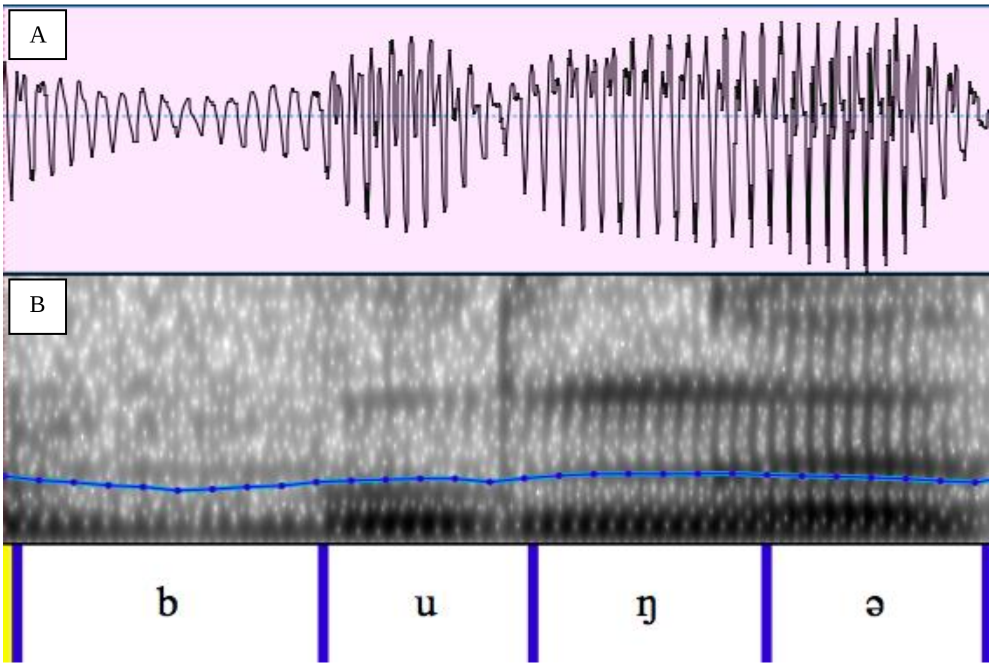


Figure 1: A time-domain representation of the acoustic waveform (Panel A) and a spectrogram analysis (Panel B) of the Malay word *bunga* or flower in English. The blue line is tracking transitions of $F_{1.}$

The intensity of normal speech conversation is about 60 dB SPL (at a distance of about 1 meter) but varies greatly with about a 28 dB difference between the lowest and the highest intensity speech sound while speaking (Denes & Pinson, 1993). Vowels tend to be higher in intensity compared to consonant sounds. The spectrum of human speech ranges from about 50 to 10000 Hz and the greatest energy is generated at low frequencies in the 100 to 600 Hz region. Male speakers generally show lower fundamental frequencies than female speakers. Spectral analysis of speech is performed by recording long sequences of connected speech that contains commonly occurring words which are later separated in smaller sequences. The summed energy for the each part of the sequence is then plotted in an intensity-frequency graph and is called the long term average speech spectrum (LTASS). Byrne et al. (1994) examined the LTASS in several languages and discovered very little difference between languages and recommended the use of a universal LTASS suitable for procedures in prescribing hearing aids and calculating the Articulation Index (AI). The AI can be defined as the degree of audibility of a speech signal, and is highly correlated with traditional speech intelligibility scores.

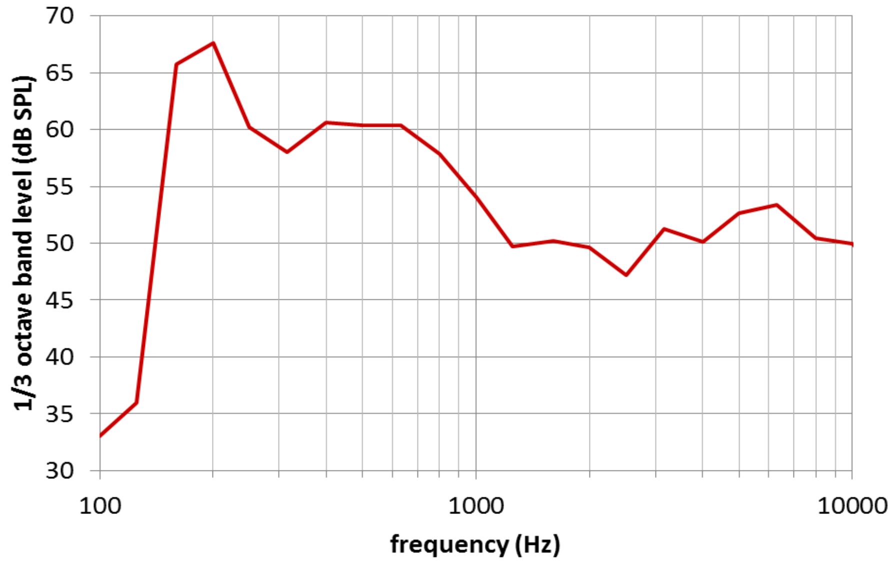


Figure 2: Long term average spectrum of speech of a female Malay speaker used in this study.

Speech is produced through several processes, namely initiation, phonation, the oro-nasal process and articulation. The initial process to produce speech can be typically described as build-up of air pressure in the lungs which is then released through the vocal cords that creates phonation. This flow of air pressure is then controlled in the vocal tract by adjusting the position of multiple articulators such as the nose, jaw, tongue, palate, teeth and lips to create either harmonics or distortion that is combined to form sounds that we hear as speech. The movements and interactions of the articulators will shape the vocal tract and restrict or release air flow making the speech sound unique to the overall structure of the tract.

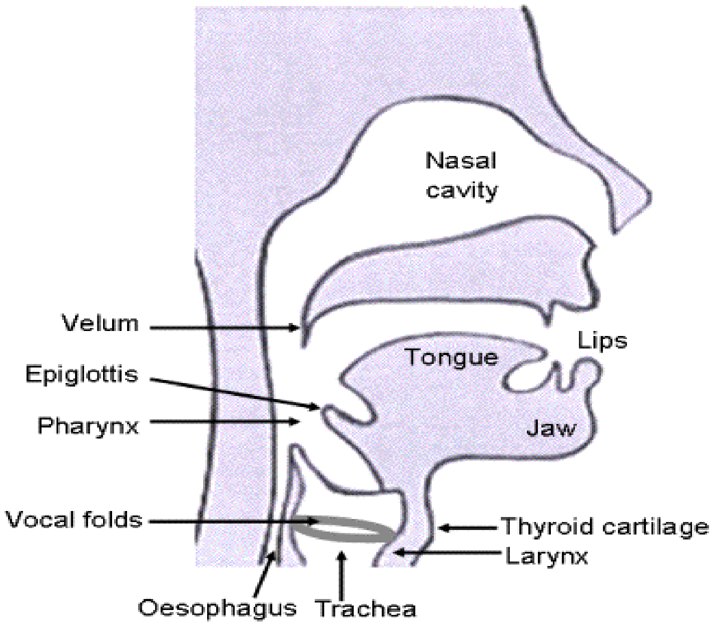


Figure 3: Organs involved in sound phonation (Image adapted from Wolfe, n.d.)

Quarter wavelength filter harmonics are created when a stable flow of air passes through an opened vocal tract and is restricted at certain points of the tract. These harmonics or resonances, which are concentration of robust acoustic peaks within the complex speech sounds, are also called formants. Phonemes of vowels sounds of English are determined by the positions of the first three formants, namely F_1, F_2 and F_3 (Lieberman & Blumstein, 1988), whereas consonants are the product of acoustic harmonics and/or distortions and are classified based on the manner and place of speech production. A summary of acoustic features of English consonants are shown in the table below:

Table 1: Types of English consonant phonemes and their acoustic features (adapted from Lieberman & Blumstein, 1988)

Type	Consonants	Acoustic production feature
Stop consonants	[p t k d k g]	Rapid release of a complete closure. Rapid release burst (5-15 milliseconds) and duration of formant transition between 20-40 milliseconds.
Nasal consonants	[m n ŋ]	Rapid release with a closure in the supralaryngeal oral cavity with an open velum. A nasal murmur occurs prior to release of closure. Murmur is dominated by low frequency sounds dominantly around 250 Hz with resonances occurring around 700 Hz at lower intensities.
Liquids and glides	[l r] liquids [w y] glides	Produced with a partial constriction in the vocal tract. Onset frequencies, duration of sound and direction of vowel transition essential to distinguish between sounds. (i.e. vowel transitions of 40 milliseconds or more will be perceived as glides, shorter sounds will be perceived as stops.
Fricatives	[f θ s v ʤ ʧ z]	Produced by partial constriction at a narrow channel by placing two articulators close together. Bursts of noise of more than 20 milliseconds with gradual onset. Presence of aperiodic noise in the spectrum.

2.1.2 Anatomy and physiology of speech perception

Speech, like any other sound is heard by a person through processes in three auditory regions, which are the external, middle and inner ear. In brief, sound travels to the peripheral ear when a certain amount of force (in this case speech production) is applied to force air particles to be displaced from their resting position. Sound is received at the external ear which channels the displaced air particles to the middle ear. Sound energy is transferred by the middle ear from a low-impedance air based medium to a higher-impedance fluid based medium which is in the cochlea of the inner ear. The cochlea senses different physical components of the speech sound and converts the mechanical stimulation to nerve impulses (n.b. these processes are discussed in more detail in later sections). At this point of the pathway, the neural signals representing the speech sounds are then sent to dorsal and ventral cochlear nuclei and the superior olivary complexes. The components of the speech sounds are segregated into different pathways where temporal and spectral analyses occur. At the levels of the lateral lemniscus, inferior colliculus, medial geniculate body and auditory cortex, the components of the sound are integrated, leading to the identification and comprehension of the acoustic cues.

In this section, a closer review is done on the hearing anatomy and function that contributes directly and indirectly to speech perception.

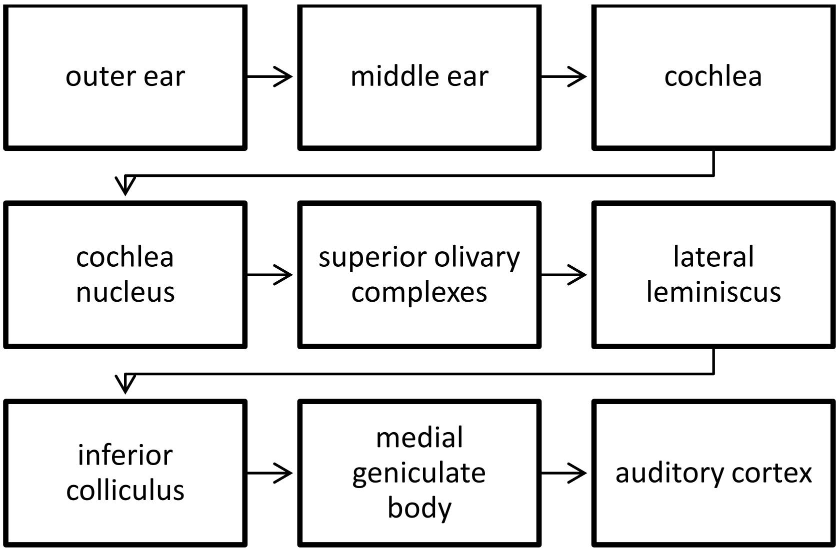


Figure 4: Simplistic diagram of major landmarks in the auditory pathway

The external ear consists of the pinna and the external auditory meatus. The pinna is characterized by structures such as the concha, tragus, lobule and helix whereas the external auditory meatus is a short, s-shaped rigid cylindrical tube which ends at the eardrum or tympanic membrane. Due to the shape of its structure, the external ear helps to amplify sound due to length-dependent harmonics that occur when sound travels through these structures (more notably the concha and the external auditory meatus). Shaw (1968) described that adults benefit from a broad fundamental resonance at about 2.7 kHz with gain of up to 15 dB contributed by the external auditory meatus that measures at about 2.5 cm. Changes in length, as observed in children, will change the peak resonance. In addition, the concha contributes to a gain of about 5 dB at the regions of 5 kHz by the same resonance principle. These two regions are especially important for speech perception as the first formants for most vowel sounds are concentrated at the region of 1.5 to 3 kHz (shown in the figure below) and softer unvoiced fricative phonemes such as /s/, / θ/ and / ʧ / are generally produced at frequencies higher than 4 kHz. Gain at these regions assists listeners in identifying and discriminating vowels and generally softer unvoiced fricatives especially in difficult listening situations.

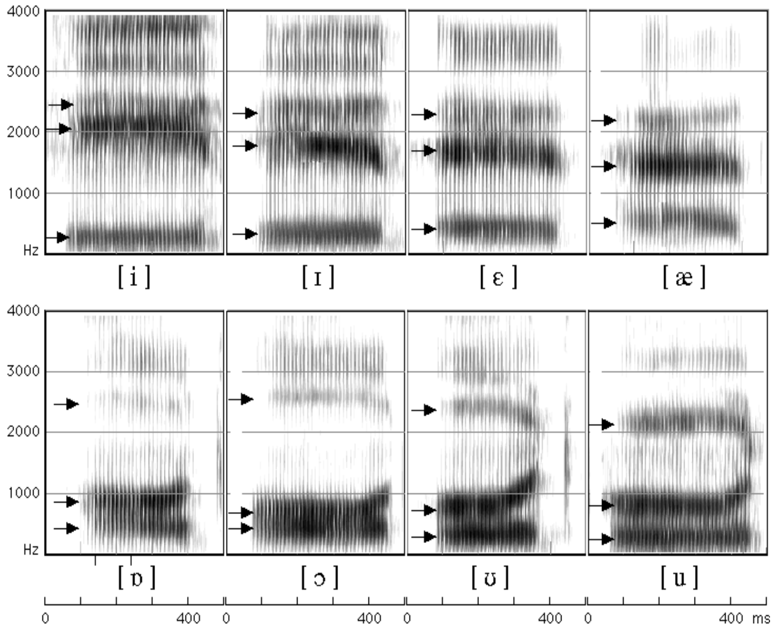


Figure 5: Spectogram of English vowels (Adapted from Ladefoged & Johnson, 2014)

One other aspect that the external ear contributes to speech perception is spatial hearing (Teranishi & Shaw, 1968) whereby the pinna assists in localization of sound, particularly in the vertical plane with a resolution of about 3 degrees.

The middle ear has two major functions: (1) to ensure energy is efficiently transferred from an air based medium to a fluid based medium through the oval window of the cochlea and (2) to perform a protective function from loud sounds by reflexively contracting middle-ear muscles that stiffen the ossicular chain. The impedance mismatch between the two media is overcome by the differences in effective areas of the eardrum and the footplate of stapes together with some contribution of the lever action and buckling effect of the tympanic membrane. Key to speech perception, this effect has caused the transmission of sound to be most efficient from 500 to 4000 Hz (Aibara, Welsh, Puria, & Goode, 2001) where most speech sounds have their energy. Huxley (1990) suggested the second main function of the human middle ear is to reduce internal physiological sounds such as chewing, breathing and even one's own speech, as these sounds would otherwise be perceived as loud at the level of the cochlea which would have a masking effect towards external sounds. This is achieved by stimulating the two efferent auditory pathways which are the middle ear muscle reflex (MEM) and the olivocochlear reflex (MOC). The first of these efferent pathways helps the muscles in the middle ear to contract, thus reducing direct bone conduction stimulation of the ossicles which in turn will reduce the amount of energy transmitted to the cochlea. The term anti-masking mechanism was used to describe how both the MEM and MOC work coherently to reduce masking effects of loud sounds by reducing the upward spread of masking by low frequency noise (Liberman & Guinan, 1998). Both mechanisms are advantageous to understanding speech by improving signal-to-noise ratio.

The inner ear consists of the cochlea and the vestibular system. The cochlea can be described as a fluid filled shell containing several vibrating membranes and large numbers of nerve fibres. Each fibre is sensitive to a specific narrow frequency band. The nerve fibres synapse with inner hair cells located along the basilar membrane, which is coiled within the cochlear shell. There are two types of hair cells, the outer hair cells (OHC) and the inner hair cells (IHC). The OHCs work as a mechanical transformer that amplifies low-level sounds whereas the IHCs release neurotransmitters to the auditory nerve. The hair cells are stimulated by different mechanisms primarily due structural differences between the cells. The OHC depends on the relative displacement between the tectorial membrane and and reticular lamina as the OHC's stereocilia are embedded in the tectorial membrane. The IHC's stereocilia are apparently free standing and deflection of the stereocilia would largely depend on the viscous force applied to them which is proportionate to the amount of basilar membrane displacement.

The ability to provide additional energy to low level sounds to enhance mechanical vibrations to specific regions of the basilar membrane is dependant on the integrity of the OHCs (Dallos, Billone, Durrant, Wang, & Raynor, 1972). This process is called the cochlea's active process or cochlear amplifier (Davis, 1983). A model of the active process (S. T. Neely & Kim, 1986; Stephen T. Neely & Kim, 1983) described it as mechanical force generators that are powered by electrochemical energy. This process allows for narrow tuning reponses within a single OHC fibre at a given point on the basilar membrane in response to a specific low intensity stimulus (see Figure 6 below). Without this active process the sharpening of tuning curve of each single OHC fibre will be lost, reducing overall sensitivity to low level sounds (Evans, 1975).

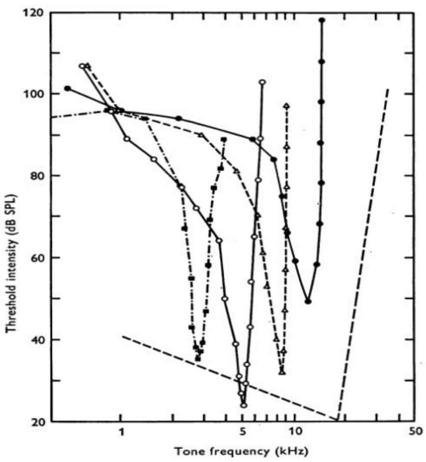


Figure 6: Response areas of cochlear nerve fibres of guinea pig. The sound pressure level of stimulus was measured at the level of tympanic membrane. (Adapted from Evans, 1967)

The IHCs require adequate mechanical stimulation to deflect hair bundles and cause ionic gates to open and shut, causing a cyclical inflow of potassium ions. This alters the voltage difference between the inside and outside of the cell and leads to the release of neurtotransmitter and stimulation of the auditory afferent fibres which in turn carry the information regarding the sound. The hair cells are organized within the cochlea to respond tonotopically, where the characteristic frequency (CF) of hair cells closer to the basal end of the cochlea are responsive to high frequency signals and hair cells contained at the apical end of the cochlea are responsive to low frequency signals. This organization called the tonotopic organization suggest an important role of place coding mechanism, where only specific regions of hair cells are excited depending on the spectral content of the signal.

In addition to this, each hair cell acts as a bandpass filter and when excited, causes action potentials and discharge of neural signals in patterns that are a function of time and intensity. The synchronization between patterns of resulting auditory nerve firing as response to the pattern of an input is called phase locking mechanism. These two mechanisms help to explain the physiology of the cochlea’s response to sound and of course speech. Speech is unique in the way that it carries large amounts of spectral and temporal information and it will stimulate large portions of hair cells within the cochlea at once, hence it is safe to assume that both mechanisms play an active role in conveying speech signals to the higher levels of structure within the auditory pathway, terminating in the auditory cortex.

While it is difficult to pinpoint the exact structure in the central auditory system that is responsible for speech perception, studies in event related potential (ERP) and functional magnetic imaging using speech stimuli have greatly improved our understanding in this field. Specific components in ERP studies such as N400, mismatch negativity (MMN), early left anterior negativity (ELAN) and P600 have been shown to have strong correlation to syntactic, semantic and phonological processes (Osterhout & Holcomb, 1993). Using specific sensory modes like audition, vision and somatosensation together with controlled presentations and recording parameters, changes within these neural generators are observable and can indicate certain processes in the central nervous system. Studies in ERP also use anomalies in the stimuli to observe changes in the components and are called error-related negativity studies. They are used to record evoked potential markers in response to auditory stimuli with unexpected speech or/and language errors (Dikker, Rabagliati, & Pylkkänen, 2009; Näätänen et al., 1997).

Lau, Phillips, & Poeppel (2008) reviewed the functional response N400 in speech processing in several ERP studies and compared it to more recent findings in functional magnetic resonance imaging (*f*MRI). They concluded that N400 is an influential marker for the extraction of words and meaning, and noted that the response it is reflected in the activity of the posterior middle temporal cortex, which is the only area to show the effects of semantic priming in *f*MRI studies across all conditions that also show an N400 effect.

In a study of MMN which is a change specific component of ERP that can be recorded without the attention of subjects, Näätänen et al. (1997) presented Finnish speakers with non-native phonemes. They found that the MMN component was significantly enhanced when the deviant stimulus were presented uggesting that MMN responses to speech sounds are language specific and associated with phonologic memory primarily in the left hemisphere of the auditory cortex. In studies of learning effects, MMN traces were also found to be associated with language learning as traces of MMN seem to be similar between non-native fluent speaker to native speakers of the Finnish language (Kujala, Tiitinen, Alku, Lehtokoski, & Ilmniemi, 1999). Additionally, changes in MMN traces were observed after brief training of non-native fricative syllables was provided to native Finnish speakers (Tamminen, Peltola, Kujala, & Näätänen, 2015). The MMN is suggested be to generated at least two supracranial processes which are the supratemporal processes (bilateral) and the pre-dominantly right frontal process (Näätänen & Kreegipuu, 2012). For interested readers, there are criticisms on ERP that should be noted, such as using presentation rate of speech are that are often not as rapid as normal speech (Rayner & Clifton, 2009) and the assumption that controlled speech stimuli such as phonemes and short sentences are representative of real-world speech perception as an unsatisfactory explanation for ERP markers (Steinhauer & Drury, 2012).

To summarize the areas in the central nervous system responsible for speech perception, evidence shows that phoneme identification could most possibly occur at around region of the superior-anterior region of the auditory cortex and early syntactic word structure for specific word categories occurs very early during speech processing, as shown in ELAN (Friederici, 2002). The N400 reflects the identification of lexical-semantic processes and is commonly used across many languages as it can be recordable regardless having the subject to be alert to the stimulus. The positivity that occurs at around 600 milliseconds is known to reflect analyses of more complicated sentences in speech and syntactic errors and is believed to originate from the frontal and centro-parietal region of the brain.

2.1.3 Theories of speech perception

In attempting to describe the complex nature of speech perception, many researchers debated on how the information in speech is extracted and processed. Critical issues that are often discussed are: (1) the ability to detect the presence or absence of speech, which often involves the discussion of the functions of the peripheral auditory system; (2) the discriminatory ability of the auditory system and memory to extract segments of sound; (3) the recognition process that involves the ability to change perceived auditory stimuli to perceptual patterns that could be recognized as speech; and (4) synthesizing abilities to create meaning out of recognized speech patterns. In this part of the literature, several popular theories of speech perception are described and discussed.

2.1.3.1 Motor theory of speech perception

Work beginning in the 1950's at the Haskin Laboratories by Alvin Liberman, Franklin Cooper and other researchers has produced an important model of speech perception called the motor theory of speech perception (MT) (Delattre, Liberman, & Cooper, 1955; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). This model has helped to describe the complex nature of speech and became the basis of the study and analysis of phonemes. It is guided by the understanding that speech perception is based on neuromotor commands called articulatory events rather than acoustical or auditory events, and it occurs when there is consistency between the produced phonemes or feature sets which are in agreement with neuromotor commands. The processes of speech production involved in MT begins with the identification of phonemes or sets of distinct articulatory feature which in turn would be changed to neuromotor commands that will give specific instruction for muscle contractions which changes the shape of the vocal tract thus producing an acoustic signal. In brief, a person's perception of speech is dependent on an internal mechanism that decodes speech in its smallest unit based on knowledge of articulatory movements. This internal mechanism called the speech decoder is hypothesized as speech sound that is analysed by an internal vocal tract synthesizer that incorporates information about the anatomical and physiological characteristics of the vocal tract and its acoustical consequences. The basis of this model can be explained by comparing the synthesized single syllable sounds with the same consonant but different vowels (e.g. /du and /di/). The rapidly changing formant transitions at the onset of each syllable carry important information. For the example below, the distinguishing element between /du/ and /di/ are the rising or falling of the onset of the second formant, and the concentrations of different levels of energy at different frequencies for both sounds. The major contributing factor that generates these physical differences is the shape of the vocal tract which is influenced by neuromotor commands.

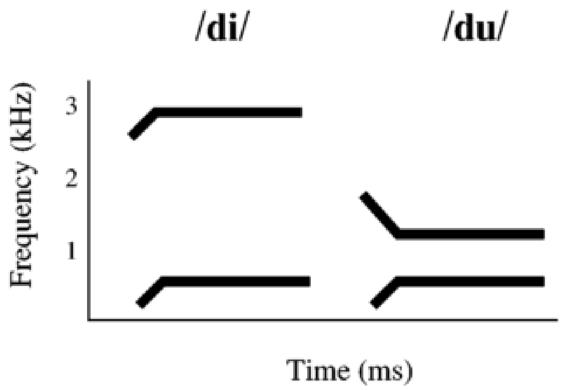


Figure 7: Formant patterns of /di/ and /du/ (adapted from Delattre, Liberman, & Cooper, 1955)

2.1.3.2 Direct realist theory

Another popular approach to understanding speech perception is the direct realist theory of speech perception (DRT) that was developed by Carol Fowler, also from the Haskin Laboratories (Fowler, 1980, 1986, 1996). Similar to MT, DRT also believes that speech perception is to related articulatory movements and not based on acoustic cues. However, what makes DRT different from MT is that it suggests perception of speech as not inherently based on articulatory events that are caused by neuromotor commands but are actual events that directly cause sensation similar to visual perception. For example, a person identifies a phoneme as a causal effect to movements of articulators rather than using internal knowledge as in MT. It implies that perception is not mediated by delays of synthesizing processes but rather the direct understanding of acoustic signals from speech production, as the signal is more than adequate to provide stimulation. In this theory, production of phonemes and other phonetic cues are coproduced and retain an independent acoustic structure together with temporal overlaps. This is rather different from MT because MT tends to suggest that speech perception is more exclusive to humans due to the intrinsic knowledge of speech production whereas DRT can be applied to a broader range of species that use acoustic signal as ways communicate (Fowler, 1996).

2.1.3.3 General auditory and learning approaches

As an alternative to MT and DRT, the general auditory and learning approaches or also known as general approach (GA) are guided by the philosophy that auditory neural or even multi-modal responses are grouped into linguistic categories. This is due to the reason that some instances of speech perception of speech stimuli matches those of nonspeech stimuli when they share similar critical temporal properties (Diehl & Kluender, 1989; Massaro, Cohen, & Oden, 1980). This model assumes that speech perception uses the same mechanisms of audition and perceptual learning of other nonspeech stimuli such as environmental sounds. In terms of speech production and perception, GA suggests that perception must precede production. This means that in order for speech sounds to be produced, the auditory and cognitive processes must be able to recognize distinctive features in the acoustic signal. This helps to explain claims of invariance of acoustic signals in speech and nonspeech sounds as well as provide links with nonhuman species.

Within the scope of GA, the TRACE model of speech perception (Mcclelland & Elman, 1986) offers some explanation of the role of memory and higher order cognitive processes. It involves the dynamic perception of three levels of unit; the feature, position sensitive phoneme and words. At any point in time, these three levels may occur, be processed simultaneously, and offer bidirectional top-down and bottom-up levels of processing. Lexical knowledge of a phrase and/or word is applied in an anticipatory pattern in time. A word is perceived when the greatest amount of activation occurs within connections of inhibitory and excitatory layers of connections. This model is commonly used in computer models to predict speech perception.

Another higher level model of speech perception that is often used in computer modelling is the fuzzy logic model of speech perception (FLMP) (Massaro et al., 1980). This model considers that there are interactions in the evaluation of acoustic features and listeners require extreme values of acoustic features for some speech sounds compared to other speech sounds. As an example, the aspiration period of noise during voice onset time (VOT) serves as an additional distinguishable feature. In this model, each syllable is presented in memory as prototype which is made up of features with specific values. Perception occurs when the signal matches the expected values which are mathematically calculated into a final score for recognized syllable. This model can apply any other external factors that might contribute to speech perception such as visual and tactile sensations.

2.1.3.4 Summary of speech perception theories

The MT showed speech perception theory depends on an internal knowledge of articulatory gestures whereas the DRT adheres to more classical sensory perception models and may provide answers to explain speech perception in nonhuman species. The GA response to these models suggests that acoustic cues play a dominant role in speech perception. Viewpoints presented in each model above indicate the complexity of speech perception. There are other speech perception theories that are not included in this literature such as the quantal theory (Stevens, 1989), and native language model (Iverson & Kuhl, 1995; Kuhl & Iverson, 1995). All serve to answer the four key elements presented in the introduction of this section of the literature which are to explain the detection, discrimination, recognition and synthesis of speech sounds. Each one supports different elements of speech perception which helps us to understand the possible mechanisms behind resolving acoustic cues into meaningful speech as well as nonspeech signals.

2.1.4 Auditory-visual speech perception

Speech perception in real-world is ultimately a multimodal phenomenon. In normal daily conversation, communication partners often can hear and see each other. This facilitates communication by providing auditory, visual and at times somatosensory cues. It is also known that being able to see the speaker also helps with language comprehension and memory of the spoken language (Campbell & Dodd, 1980).

One of the pioneering works in auditory and visual influences in speech perception was by McGurk & MacDonald (1976). They initially observed that normal hearing listeners reported hearing the syllable /da/ when they were presented with a video of a woman saying the syllable /ba/ but was dubbed with the sound /ga/. This prompted them to construct a more structured experiment to explain the influences of visual input in speech perception. They performed experiments in auditory and auditory-visual modes on subjects of various age groups. Subjects were required to observe and/or listen and report the sound they perceived by looking at recordings of a woman whose voice has been dubbed to create four videos of voice-to-lips mismatch. Considerations were taken to ensure that the voice onset time was the same. They reported that in the auditory only condition, accuracy was high for all subjects at over 90% correct responses. More errors and varied responses were reported in the auditory-visual condition as accuracy dropped by 7 to 35 percent and children showed higher dependency towards visual stimuli compared to adults. The response that was observed earlier was also reported where listeners gave a response which was a combination of the two modalities and was transformed into a new element originally not present in either the auditory or the visual recordings. This ‘fused’ response in now commonly known as the “McGurk effect” and is widely used in other studies of auditory-visual speech perception. They concluded by noting that auditory-based theories of speech perception do not provide adequate explanation of their findings. Using *f*MRI, Jones & Callan (2003) used modified versions of the McGurk & MacDonald experiment to investigate the relationship between brain activation and the degree of auditory-visual integration during a speech perception task. In the auditory-visual task, the audio stimuli were either in total synchrony or 400 milliseconds out of phase with the visual stimuli. They reported that active brain regions during this task include the superior temporal and inferior frontal gyrus as well as the extrastriate, premotor and posterior parietal cortex. The McGurk effect was found to correlate positively with the activation of the left occipito-temporal junction which is often associated with processing visual motion. Using this finding, they proposed that auditory information modulates visual processing in auditory-visual speech perception.

According to Grant, Walden, and Seitz (1998), modality must be considered in an auditory-visual communication environment. They suggested a framework with bottom-up information extraction, integration and a top-down linguistic process similar to the motor theory of speech perception but also included visual cues as a key component. To test this framework, they studied auditory-visual integration abilities of hearing impaired subjects using nonsense syllables and sentences to explore relationship between auditory, visual and auditory-visual cues separately. Auditory-visual integration is an element to measure the ability to utilize both modalities simultaneously without relative contributions from unimodal auditory only and visual only encoding. This is done by making predictions of auditory-visual performances based on speech feature recognition or a matrix of confusions obtained from a separate modality. First, Grant et al. found significant improvements when both auditory and visual cues were presented simultaneously compared to the auditory only condition. Secondly, when presented with auditory cues alone, voicing and manner of speech played a more prominent role in speech recognition whereas place of articulation provided more speech recognition cues when visual inputs were presented by itself. This indicates the interdependence of both types of sensory input in gathering as much information as possible in understanding speech. They also found that the ability to use both auditory and visual cues varied across subjects with those with higher auditory-visual integration getting more auditory-visual benefit.

To identify the influence of age in auditory-visual speech perception, Tye-Murray et al. (2008) investigated the abilities of adult subjects with different age groups using a test developed in-house called the build-a-sentence test (BAS). Participants were screened for hearing impairment, speech abilities and history of central nervous system disorders. In this experiment, the BAS test was presented to participants in favourable and unfavourable visual and auditory conditions. They found that the performance of older adults deteriorated more compared to young adults when the viewing and listening conditions were unfavourable. This is contributed to by age as older adults experience degradation in their auditory, visual abilities and for some, their cognitive abilities as well. To examine age and hearing impairment as factors that affect auditory-visual speech perception, Tye-Murray, Sommers, & Spehar (2007) conducted experiments in normal and impaired hearing older adults in auditory only (A), visual only (V) and auditory-visual (AV) conditions using speech materials in the form of consonants, words and sentences. They also assessed the integration enhancement (IE) abilities of participants, which is a measure adapted from auditory-visual integration that uses a simple probability matrix. The working hypothesis for this study was that hearing impaired listeners could develop better lipreading skills and auditory-visual integration, as the gradual onset of hearing impairment would increase reliance on visual cues of speech and that would train them to be better lip readers. A, V and AV responses were obtained from participants using consonants, words and sentences, and revealed that the hearing impaired participants performed significantly better than their normal hearing counterpart in the V condition for the word test but no differences were found for the consonant and sentence format of the experiment. This indicates that performance of lipreading is highly individual and not related entire to age and hearing impairment. No statistical differences were found in the integration enhancement measure in both groups suggesting hearing impairment is not a factor in a person’s ability use both auditory and visual information in speech perception. Blamey, Cowan, Alcantara, Whitford, & Graeme (1989) studied the same effect including tactile (T) in combination with A, V and AV conditions. Four normal hearing subjects were trained and tested and results showed that the addition of the tactile input produced significant improvements but were marginal compared to auditory and visual aids. This study suggests that audition and vision are key components in speech perception and tactile cues could be beneficial when hearing and/or vision is limited. Based on the studies above (Blamey et al., 1989; Grant et al., 1998; Tye-Murray et al., 2007a; Tye-Murray, Sommers, & Spehar, 2007c; Tye-Murray, 1992), we can conclude that: (1) the ability to lip read is independent of age and hearing levels; (2) hearing impairment affects auditory-visual speech perception and improving audition could improve speech perception; (3) the ability to integrate both auditory and visual cues is independent of age and hearing levels; (4) tactile information could assist in speech perception in any conditions with training; and (5) investigating AV integration abilities is important clinically as listeners with good integration abilities make better speech readers and possibly have better outcomes with intervention.

Therefore, in order for clinicians to establish true communication abilities, a speech test should incorporate auditory and visual inputs, as it is important to categorically identify the specific aspects of particular modalities or abilities that could be intervened or investigated further. A limitation that is described by the tests used in the studies above is that the AV materials were either in the form of syllables or words, and the sentences were constructed from daily sentences. These tests are helpful in some situations but are not helpful in audiology clinics where clients are required to be tested repeatedly without memorization and significant floor or ceiling effects. To achieve this, Trounson and O'Beirne (O'Beirne, Trounson, McClelland, & Maclagan, 2015; Trounson & O'Beirne, 2012) developed a matrix sentence test that uses both auditory-visual cues and adaptive measures to ensure a test with high redundancy and low probability as well as no floor or ceiling effects. This seems to be a natural evolution for a speech test as speech understanding is not only limited to auditory cues but also involves an individual's ability to recognize speech patterns through vision. Both auditory and visual information are used concomitantly to formulate meaningful interpretations of sounds using their intrinsic linguistic knowledge. This is important as vowel sounds are not that visible, but are easily heard, but consonants are not easily heard in noise but are more visible due to the nature of speech production. For this study, a Malay matrix sentence test that assesses the auditory, vision and integration of both types of inputs was proposed.

2.2 Introduction to speech tests

Evaluation of speech perception is mainly achieved in audiology clinics through the use of speech audiometry. Speech testing also provides support to the inadequacy of pure tone audiometry in measuring real-world communication disability (Killion & Niquette, 2000) even though it was suggested that there were some mathematical correlations between pure tone results and speech test results (Fletcher, 1950). A broad definition of speech audiometry is that it is a technique to measure some aspect of hearing ability using standardized samples of a language that are presented methodically (Carhart, 1952). Speech audiometry was originally developed out of the work conducted at Bell Telephone Company from the 1920s to 1930s to evaluate the efficiency of communication systems (Fletcher & Steinberg, 1947) and interest increased after World War II as soldiers returned with hearing loss. During that period, there was an immediate need to examine the effects of hearing loss on speech perception as well as the experimentation on the effectiveness of aural rehabilitative regime. Speech tests have now been utilized as a routine part of the audiological test battery for many years and are integral to gauge the extent of a hearing disability.

A speech intelligibility test usually measures the speech-reception threshold (SRT), which is the lowest level at which a subject is able to correctly discriminate 50% of the speech material. Other parameters that are observed in speech audiometry are the speech detection threshold (SDT), the point of maximum score and the slope of intelligibility. A speech test with a steep curve and a measurement error of about 1 dB should be able to reliably differentiate between normal hearing and hearing impaired listeners (Suter, 1985). Speech tests nowadays are designed to assess certain listening abilities with tests varying in different aspects of their composition. To effectively achieve this objective, considerations are given to the target group (adult or children), classes of speech materials (monosyllabic words, disyllabic words or sentences), presentation modes (in quiet or with competing noise and/or using fixed or adaptive measurement techniques) and type of response required and/or observed. Some researchers have chosen to design speech-in-noise tests to be presented with fixed signal-to-noise ratio (SNR) while others have used an adaptive SNR approach. Examples of speech tests using fixed SNRs are the Connected Speech Test and the Speech Perception in Noise Test (SPIN) (Kalikow, Stevens, & Elliott, 1977) , whereas examples of adaptive measure include the Hearing in Noise Test (HINT), which has been produced in many languages (Quar et al., 2008; Vaillancourt et al., 2008; Wong, Liu, & Han, 2008) and the Digit Triplet Test (DTT), which has also been adapted for many languages (Smits et al., 2004; Wilson, Burks, & Weakley, 2005; Zokoll et al., 2012). Some speech tests may adopt a phonemically balanced equivalent list approach which indicates that it represents a proportional representation of sounds that occur in everyday speech. The reason for using this is that if a listener is unable to perceive a commonly used phoneme, the level of handicap would be larger than it would have had a phoneme been a less frequently occurring one. However, the use of a phonemically balanced speech test material to predict real word communication is questionable since there are many other qualities in speech that are important such as temporal cues and the spectral changes during the transition of one speech sound to another (Dillon & Ching, 1995). The table below shows a summary of many available speech tests with different designs.

Table 2: List of speech tests with different designs (source: Dillon & Ching, 1995)

Test Name	**Type of signal**	**Target group**	**Response type**	**Ability tested**
AB word test (Boothroyd, 1968)	Monosyllabic words using CVC words in isolation	adults	Open set, repeat words	Phoneme/word identification
CUNY Nonsense Syllable Test (Levitt & Resnick, 1978)	Nonsense syllables either using CV or VC.	adults	Closed set with 7 to 9 choices, mark on response sheet	Phoneme identification
Kendall toy test (Kendall, 1953)	Familiar monosyllabic words	children (>3 years old)	Closed set with 10 choices, point to toys	Word identification
CID Auditory test W1 (Hirsh et	Familiar spondee words	adults	Open set, repeat words	Word identification

al., 1952)				
HINT (Nilsson et al., 1994)	Everyday sentences with high predictability	adults	Open set, repeat key words	Word identification
SPIN (Kalikow et al., 1977)	Sentences containing high and low predictable words, varying word length (5-8 words)	adults	Closed set, repeat last word in sentence	Word identification
Digit triplet test (DTT) (Smits et al., 2004)	Three digits in random sequences	adults	Closed set, press numbers/digits	Digit identification, mainly used as screening tool
Matrix sentence test (Hagerman, 1982)	Semantically unpredictable sentence in fixed grammatical structure	adults	Closed set, either repeat or select words	Word identification
Australian adaptation (Plant, Phillips, & Tsembis, 1982)	Questions related to self, and familiar issues	adults	Open set questions	Understanding and answering questions

Previously most speech tests only used monosyllabic words presented in quiet in free field or under headphones, and were used to either evaluate the validity of the audiogram or to assist in the hearing aid selection process by comparing aided and unaided responses (Dillon, 2000). Attempts to ensure the relevance of speech tests in assessing real-world communication have led to the discussion of the applicability of presenting monosyllabic materials in quiet, as difficulty in understanding speech in noise is considered as the greatest handicap associated with hearing impairment (Smits, Kramer, et al., 2006; Walden, Schwartz, Williams, Holum-hardegen, & Crowley, 1983). Killion & Niquette (2000) suggested that speech in noise tests should be used as a compulsory routine assessment as they found poor correlation between results of single syllable speech tests in quiet and a patient's ability to hear in noise. They coined the term "SNR loss" to highlight this specialized component of hearing. A review on the variation between pure tone audiometry and speech in noise test scores and how to apply SNR loss is found in Taylor (2003). With this considered, today there are many speech tests that use sentences in quiet and noise, and these are widely used in clinics for many applications.

As a tool for audiological investigation, most speech tests are used to cross check the pure tone assessment of hearing loss (Fletcher, 1950). The SRT of individual ear is usually compared to the average pure tone results of the same ear. This could be very valuable especially when a pseudohypoacusis is suspected. Discrepancies between pure tone and speech results could give some information to the audiologist about whether to proceed with an objective type test is necessary. Speech audiometry is also used for differential diagnosis of hearing loss. In the AB word test for example, if speech perception performance in a normal hearing listener deteriorates after reaching a maximum by a significant amount as test intensity is increased, an audiologist could suspect a hearing loss that is possibly neural rather than cochlear in origin and further investigation using auditory evoked potentials or cochlear emissions may be warranted.

One area of investigation where speech tests play an important part is for the diagnosis of central auditory processing disorders. Speech tests are used either through monotic or dichotic presentation. Monotic or monaural testing is conducted using earphones, with the signal heard in one ear at a time. Dichotic testing is conducted by presenting different acoustic stimuli with simultaneous onset and offset times by earphone to the two ears. Dichotic stimuli include consonant-vowel syllables, digits, monosyllable words, two syllable words (spondees), and sentences (Bellis & Ferre, 1999; Mukari et al., 2006). Subjects with auditory neuropathy-like features could also benefit from tests using speech in noise. For example, a study by Rance et al. (2010) showed significantly poorer speech perception results in subjects with Freidreich's

Ataxia (FA) than their sensorineural counterparts. They suggested the inclusion of speech perception in noise as part of the investigation of patients suspected with auditory neuropathy/dys-synchrony spectrum.

Speech tests are also used as screening tools for hearing impairment. The Kendall toy test (Kendall, 1953) is useful in assessing preschool children as the speech materials are familiar items and could be easily conducted even on children with poor expressive abilities. This allows children to be screened without the use of pure tones or an elaborate audiological setup. The use of speech test as a screening tool is not limited to children. The digit triplet test (DTT) (Jansen, Luts, Wagener, Frachet, & Wouters, 2010; Ozimek, Kutzner, Sęk, & Wicher, 2009; Smits et al., 2004) is now used in many European countries as part of the HEARCom initiative to provide a self-administered hearing screening opportunity to the public (Zokoll et al., 2012). Smits, Kramer, et al. (2006) compared the effectiveness of DTT through telephone to the use of five questionnaires on self-perceived hearing impairment and found that the questionnaires could only correctly account for 69% of subjects with hearing loss compared to the DTT which correctly accounted for 92% of subjects. This suggests that an objective evaluation such as DTT has a higher level of sensitivity in identifying hearing impairment compared to self-perceived hearing loss.

A major application for speech tests is in the evaluation of the effectiveness of amplification. One method that it is commonly used for is to compare aided and unaided responses. For this purpose, the level of speech and noise and type of noise are important (Dillon & Ching, 1995). Information from this type of testing provides audiologists with an objective counselling tool to help the patient understand the benefits of amplification as well as examining the improvement in speech perception. In amplification, another application that speech tests are used for is to evaluate the effectiveness of hearing aid electroacoustics such as compression, noise reduction and gain response. Because there are many parameters to be compared and testing needs to be done numerous times, the speech test must be able to provide materials with equivalent intelligibility rating. For this purpose, a sufficient amount of speech materials is required to ensure equal assessments of the many aspects of hearing aid electroacoustics and at the same time making sure that the listener's response is not through memorization.

2.3 Factors in the design of speech tests that influence measurement outcome.

2.3.1 Speech material

The process of considering suitable materials should involve the selection of highly familiar words to ensure listeners have minimum dependence on their vocabulary and phonological knowledge or speech production skills. Further analysis of the suitable materials can be done by assigning materials to a number of test sets that would be equal in difficulty perceptively. Miller, Heise, & Lighten (1950) discussed the importance of context and redundancy in test materials. Speech tests containing contextual information (such as those using sentences or nonsense syllables) have shallower psychometric curves compared to speech tests without contextual information, such as digit tests. This is explained by learning effects and the listeners understanding of topical, lexical and phonological constraints. They also summarized the importance of the set size by suggesting that the amount of acoustic information should be proportional to the number of forced choices, i.e. less acoustic information for a small number of alternatives. This could also explain the training effect found by Wagener et al. (2003) in the Danish matrix sentence test. In this study, subjects were required to listen and repeat sentences that had the same structure and to avoid memorization; the words used were randomly changed. In their study, the SRT values significantly improved after two list trials and remained relatively constant after the third trial. Here, a training effect was found after better familiarization with the test materials and protocols which can be directly contributed to the high number of words used. Hence they recommended that a training session is conducted before any measurements are taken in order to reduce this effect.

An important factor to consider when designing a speech test is the method of selection of sentence materials and the speaker used in the recording. Versfeld, Daalder, Festen, & Houtgast (2000) compiled language samples from large databases which were later vetted to ensure their social relevance and validity. By carefully selecting sentences which contained a selection of five words and also applying adaptive measure to estimate SRTn, a higher slope of intelligibility was achieved.

They identified that test accuracy was enhanced by increasing the number of independent items that could be remembered by listeners.

2.3.2 Optimization procedure

The aim of this process is to ensure all speech materials are homogenous in terms of their difficulty using the principles of psychometric theory. It normally involves compiling equally sized and balanced (phoneme, word or sentence distribution) speech test materials and evaluating them in normal hearing listeners. Terms such as optimization, normalisation or equalization are used inter-changeably between researchers but these terms usually have the same aim. The importance of normalisation is highlighted in Mendel (2008). Plomp & Mimpen (1979) developed a new speech in noise test using this method by evaluating 170 sentences at 50 dB A. They level adjusted sentences using specific set of rules in a two tier optimization technique. Level adjustments were kept under 2 dB and items that required adjustments more than this level were removed from the test which left a sum of 130 of optimized sentences in this test. As the result of these adjustments, the test showed a standard deviation of measurements of 0.9 dB between lists. The slope of intelligibility was comparable between sentences and each sentence had steep slopes, indicating that it had been optimized for intelligibility. Warzybok, Hochmuth, Laan, & Kollmeier (2005) compared the advantages gained in the improvement of slope scores after optimization for the Russian, Polish, Spanish and German matrix sentence tests and found an average improvement of 4%/dB. This indicates that the tests showed better sensitivity after the normalisation process which is essential to produce reliable results and minimize any errors due to test construction.

2.3.3 Tracking procedure

One consideration that contributes to improved efficiency is the method used to estimate the SRT in sentence intelligibility tests. Plomp & Mimpen (1979) described an adaptive “up and down” or staircase procedure to measure SRT by initially presenting the list repeatedly by increasing sound level until subject is able to reproduce the sentence correctly. At this point, the sound pressure level is reduced and/or increased (using 2 dB steps) depending on the ability of the subject to respond correctly. The steps are repeated for all of the test materials. Plomp and Mimpen tested the reliability of this method in 5 different conditions with each condition tested twice.

The result of their investigation showed the standard deviation of individual SRT values for both tests was 0.9 dB and a distinctly narrower (15%/dB) psychometric curve as compared to some other previous studies.

Brand & Kollmeier (2002) described a different adaptive procedure using systematic simulations to simultaneously estimate the threshold and slope of speech sentence tests by decreasing step sizes. They found that it is possible to achieve a reliable level of speech reception threshold with a standard deviation of 1 dB and slope estimate of 20% to 30% using at least 30 sentence materials using a word scoring system. Another method of adaptive estimation of SRT using the maximum likelihood procedure (MLP) using statistical simulations is by Zera (2004). This procedure estimates the response shape of the psychometric curve through known normative investigation to enable faster identification of multiple points in the subject's response. To compare which method is more efficient, Versfeld et al. (2000) used their finalized list and compared results obtained using the simple averaging methods by Plomp & Mimpen (1979) and the MLP and concluded that the method of simple averaging seem to be as effective as the more complex maximum likelihood procedure. Sincock (2008) found the MLP showed advantages over the adaptive staircase method in terms of processing time in speech audiometry but it showed poorer reliability. Regardless of the technique in adaptive testing, the potential of an adaptive measure is that it is highly efficient and reliable regardless of the manipulation of either signal or noise (Wagener & Brand, 2005).

2.3.4 Speaker selection

Mullennix, Pisoni, & Martin (1989) studied the effects of multiple speakers in word recognition tests and found that the effects of speakers to be highly significant with subjects scoring differently should the same speaker change his or her vocal effort. Word materials or lexical rules were examined and were found to have no significant effect on the experiment. They concluded that speaker selection is crucial in test construction especially when differences are more apparent in difficult listening situations. Precautions should be made to ensure speech stimuli remain consistent throughout the test especially when designing a speech test using multiple speakers as between speaker responses will be markedly different.

Bradlow, Torretta, & Pisoni (1996) used a multi-speaker database containing intelligibility scores of 20 different speakers and investigated the effects of gender, fundamental frequency and speaking rate. After discovering that fundamental frequencies of speakers and speaking rate had no effect on intelligibility, they looked to specific acoustic-phonetic properties of the speakers and found speaker specific differences such as vowel spaces and voicing between speakers may have significant influence on intelligibility. In terms of gender differences, female speakers were found to show a greater range of fundamental frequency compared to male and using female speakers seemed to improve test intelligibility in listeners compared to male speakers. As a general recommendation for the matrix sentence tests, the use of a female speaker is encouraged as it is a compromise between the voice characteristics of a male and child speaker (Akeroyd et al., 2015).

Versfeld et al. (2000) also found significant differences between subject responses to two male and two female speakers. The differences varied across test materials despite a controlled recording procedure to ensure inter-speaker variables were fixed as much as possible. Analysis of variance showed significant interaction between lists and speakers. They attributed the differences in result to speaker's articulation and intonation within sentences. A similar conclusion was also made for the matrix sentence tests when multi speaker analysis was done (Hochmuth, Jürgens, Brand, & Kollmeier, 2015). It was agreed that speaker differences was the main influence in the differences in intelligibility between matrix sentence tests of various languages and test design and language did not significantly affect intelligibility for the matrix tests.

2.3.5 Method of response

As mentioned in the previous section, some tests are specifically designed for the paediatric population. One major consideration for the paediatric population is the method of response for a speech test. Most speech tests for children adopt a closed set response such as picture pointing tasks as it would be less demanding for them. Clopper, Pisoni, & Tierney (2006) investigated the effects of open and closed set response and hypothesized that results would be better in closed set tasks as an open set task would require a higher cognitive demand. Their results suggest a closed set response uses phonological and lexical constraints (top-down) whereas open set tasks demand acoustic-phonetic activation and lexical memory (bottom-up) because a listener cannot use the process of elimination. Both types of response have their limitations; closed set responses are prone to training effects and guessing bias (Ozimek, Kutzner, & Libiszewski, 2012) whereas open set responses are not appropriate for listeners with disordered speech or limited language abilities. A general approach that can be applied is to adopt both methods to reduce error.

2.3.6 Masking noise

Speech understanding in noisy environments has been found to be the most difficult task for hearing impaired listeners (Kramer, Kapteyn, & Festen, 1998). As hearing professionals, it is important to understand how noise would affect listeners and determine possible solutions and for this reason, we should understand how different types of noise affect overall listening performance. It is known that the type of masker plays a huge factor in the intelligibility of the test. MacPherson & Akeroyd (2014) conducted a systematic review of 139 studies related to speech in noise tests. From this collection of studies, 885 individual psychometric slopes were fitted with a common logistical function from which the slope estimation was produced. Values of slopes were different between speech tests from as low as 1%/dB to as high as 44%/dB. Two key elements of the masker were identified to have a significant influence in slope of intelligibility which is the type of masking noise and number of maskers used within tests. For example, speech-type maskers resulted in shallower psychometric functions compared to amplitude noise masker or static maskers. This was consistent with findings in Hochmuth, Jürgens, et al. (2015) and Van Engen, Phelps, Smiljanic, & Chandrasekaran (2014). Increasing the number of maskers also increases the steepness of slope, with an average improvement of 4%/dB by increasing from one to two maskers. This is because by using one competing speaker, listeners may benefit from "release from masking" and take advantage of acoustic and linguistic cues from target stimuli.

To further discuss issues regarding masking noise in speech tests, definitions of informational masking and energetic masking are described below. It is important to note that there are controversies regarding its definition of these mechanisms however it is not within the scope this review to discuss this. Additional reading regarding this matter can be found in (Durlach et al., 2003; Lutfi, 1990; Pollack, 1975;

Watson, 2005). For the purpose of this review, energetic masking refers to masking that occurs in the auditory periphery with aim to reduce intelligibility by making target stimuli inaudible or less audible, specifically by disrupting speech-related patterns of vibration on the basilar membrane. An example of this is the use of test specific noise in the matrix sentence tests. Informational masking refers to the use of a masking noise that interferes with speech reception at higher levels of auditory and cognitive processing. An example of this is the use of multitalker babble noise in the matrix sentence tests that reduced SRTn scores due to competing attention, increased cognitive load and linguistic interference.

Energetic masking is commonly seen in speech tests such as the HINT, DTT and MST using speech-shaped noises that match the speech spectrum of the stimuli. It is viewed as the optimum background noise because of its ability to produce reliable responses (Schädler, Warzybok, Hochmuth, & Kollmeier, 2015; Wagener & Brand, 2005). Using speech-shaped noise eliminates acoustic and linguistic cues of the target stimuli therefore listeners are mostly dependent on their bottom-up hierarchy of information processing ability to listen in noise with minimum influence of language and test specific (Hochmuth, Jürgens, et al., 2015). Informational masking is commonly used in speech tests to assess listeners' ability to resolve a speech signal within competing speech noise (such as 1 or 2 speaker competing speech or a multitalker babble). Due to contextual redundancy, using informational masking will be detrimental to SRTn when compared to energetic masking as more attention is needed for the listener to capture speech input. Lutfi (1990) suggested a mathematical estimate of how much masking is needed to be regarded as informational masking. He reviewed existing data from other studies within the scope of informational masking and showed possible interactions between the number of maskers and the percentage of detrimental effects on speech tests. He noted that the model is oversimplified and it is important to know that the relative amount of informational masking will be different from one test to the other.

In this study for the Malay matrix sentence test, test specificand 6-talker babble noises were generated as competing noises to produce energetic and informational masking effect and for listeners at varying hearing levels.

2.4 Sensorineural hearing loss and speech perception

It is well understood that hearing impaired listeners have difficulty listening in challenging situations like places that are noisy and highly reverberant. People with sensorineural hearing loss show difficulties in listening that are proportional to the degree of their loss. This means a person with an average hearing threshold of around 30 dB HL is likely to have fewer difficulties in speech understanding as compared to a person with an average hearing threshold level of 60 dB HL. People with profound hearing loss will generally have problems in listening to speech in all types of listening environment. In this section of the literature, we look at issues of audibility, intelligibility and the ability to resolve speech information in sensorineural hearing loss listeners.

The differences in the ability to hear in noise in sensorineural hearing loss patients can be viewed from many points of discussion. Psychophysical studies have allowed better understanding of signal processing in normal and impaired hearing people. For example, Moore & Glasberg (2004) described the differences between the models of the auditory filters between normal and impaired hearing subjects. Auditory filters, which are approximations of the cochlear response to a specific sound on the basilar membrane (BM), were found to be consistent between normal hearing subjects whereas filter shapes were considerably different from one subject to another in impaired hearing subjects. In subjects with unilateral hearing loss, the impaired ear showed a flatter auditory filter shape as compared to the normal hearing ear. The differences in perception could be due to the loss of the outer hair cell (OHC) active mechanism or damage to the inner hair cells (IHCs). The loss of OHCs and IHCs can result in losses in both sensitivity and clarity. The loss of clarity and sensitivity would need to be compensated by significantly higher SNR for more favourable listening as compared to their normal hearing counterparts (Dillon, 2000; Moore, 1996). The OHCs are known to contribute to an active process where soft sounds are amplified by providing additional mechanical force at the basilar membrane (S. T. Neely & Kim, 1986). OHCs are more vulnerable than IHCs, and their damage is commonly linked to noise exposure, ototoxic medication and infections. Listeners with OHC damage are unable to respond to low level sounds as the ability to amplify them is impaired or lost. In an experiment using guinea pigs, the loss of gain of this process could contribute to as much as 55 dB (Patuzzi, Yates, & Johnstone, 1989) which is also consistent with the findings in distortion-product otoacoustic emissions (DPOAE) where, only subjects with hearing threshold levels of not more than 60 dB HL were able to exhibit cochlear emissions (Harris, 1990; Moulin, Bera, & Collet, 1994). As speech sounds are characterized by harmonics or noise at different frequencies and intensity levels, listeners with mild sensorineural hearing loss may not have huge difficulties in speech perception. This was found to be true in experiments where speech perception of subjects with mild hearing loss were similar to normal hearing ones. Humes, Dirks, Bell, & Kincaid (1987) studied four hearing impaired listeners and twelve normal hearing listeners. The normal hearing listeners were grouped into four equal groups and were masked to simulate the hearing impaired listeners' ability to hear in quiet. They found that two hearing impaired listeners scored better results than their masked normal hearing counterparts whereas the other two hearing impaired listeners scored equally. Similar findings was described in Zurek & Delhorne (1987) where they found that after compensating for issues of audibility by providing sufficient loudness to subjects with flat and sloping mild to moderate hearing loss, similar scores were obtained between the normal and impaired hearing groups. In term of listening to speech, this would not represent a huge problem as the redundancy in speech information will assist in the speech perception. For example, vowel sounds contain large spectral information which is gathered from different regions of the cochlea and, together with inherent phonological rules, this information can be used to compensate for the effects of reduced audibility. By increasing the loudness of sound either by using hearing aids or other assistive listening devices, most of the issues with reduced audibility can be overcome.

In moderate to severe hearing losses where possibly both the OHCs and IHCs are damaged, amplification is needed for soft sounds to enable the basilar membrane vibration reach the threshold for neural activation. This presents different levels of challenge for signal and speech processing where there is loss of sensitivity to loudness as well as frequency selectivity and resolution. When the hair cell tuning curve is lost, there will be a broadening of response at the BM making frequency selectivity particularly difficult at low levels (Moore, 1996). The second complication that may arise in this situation is off-frequency listening where signals are identified by hair cells which are bordering the damaged cells. In the situation where the IHC is non-functioning, transduction will not occur within that region and it is therefore referred to as a ‘dead region’ (Moore, 2001). It is known that dead regions are not uncommon in listeners with moderate to severe hearing losses. Consideration needs to be made for amplification using hearing aids as information is not sent within the range of characteristic frequency (CF) of the IHC immediately adjacent to the dead region. In Moore (2001), a review on speech perception abilities for listeners with dead regions at low, high and mid frequencies were made. He described that listeners with dead regions of the cochlea are more susceptible to off-frequency listening hence in broadband speech sounds, a person with low frequency dead regions may retrieve very little or no information about low frequency sounds from IHCs that are tuned to medium to high frequencies. Similar findings were reported in studies in subjects with high frequency dead regions, however listeners with mid frequency dead regions were found to he able to use information from the low and high frequency regions, and therefore reported minimal effects on speech perception ability. Additionally, amplification at the dead regions either made no difference to speech perception or caused it to be worse compared to no amplification at all. It was suggested that listeners with basal end dead regions (high frequencies) should use hearing aids with frequency compression as some benefit was found if amplification was provided one octave above the edge frequency of a dead region. A clinical tool to detect dead regions of the cochlea is the threshold-equalizing noise (TEN) test (Moore, Huss, Vickers, Glasberg, & Alcántara, 2000). Using the information above, studies of dead regions of the cochlea, and psychoacoustic studies measuring temporal resolution in sensorineural hearing loss (Dreschler & Plomp, 1985; Glasberg & Moore, 1989; Peters, Moore, & Baer, 1998), we can conclude that listeners with mild hearing loss will primarily have issues with audibility and listeners with moderate to severe hearing loss would have issues in both audibility and clarity.

For listeners with more severe hearing losses, problems with audibility are exacerbated by issues of intelligibility even at suprathreshold levels. This is because inadequate information can be gathered and transduced by the damaged IHCs and neurons and providing amplification may do very little to improve speech perception. Dreschler & Plomp (1980) studied the relationship between several psychophysical studies and speech perception tests in ten sensorineural hearing loss listeners. They found that phoneme perception studies could not explain speech perception abilities; hence the ability to understand speech does not primarily depend on phoneme detection but is contributed to by other parameters. They also explained that phoneme perception was more dependent on the configuration of the audiogram rather than the mean hearing levels.

Speech tests are designed to quantify the ability to hear speech sounds in either quiet or in noise. Taking considerations of the many arguments put by previous authors (e.g. Kalikow et al., 1977; Killion, 1997; Nilsson et al., 1994), it would be reasonable to use meaningful sentences as they have better face validity in simulating real-world hearing performance. This is because sentences contain all the linguistics elements of daily communication and also require the working auditory memory required in listening. Killion & Niquette (2000) proposed including speech-in-noise tests as part the of audiological test battery in their review of several speech-in-noise tests in normal and sensorineural hearing loss subjects. They concluded that the ability to hear in quiet is almost entirely independent of the ability to hear in noise. Speech tests commonly incorporate a speech-shaped noise or a temporally modulated noise as part of their design. The speech-shaped noise is used so that the speech information can be effectively masked as the noise has the same long-term average speech spectrum as the speech. Fluctuating noises are used to replicate noises that are associated with common competing sounds that are heard during conversations. These two types of noise will produce different outcomes for sensorineural hearing loss listeners.

In speech tests using speech-shaped noise, SRT levels of sensorineural hearing loss listeners are more predictable, as the entire speech stinmuli is masked and performance are primarily dependent on the level of audibility. In an experiment using two types of speech-shaped noise and the Oldenburg matrix sentence test, Wagener & Brand (2005) described no statistical difference between SRT for the two noises in subjects with varying levels of hearing. As both types of noise had masked the speech signal entirely and were comparable in term of their long-term spectra, they had no effect on the outcome of the test. In the same study, they conducted a test-retest evaluation and found larger variations in SRT when fluctuating noise was used, indicating better consistency when using speech-shaped noise. As for fluctuating noises, Plomp (1994) described the difficulty faced by hearing impaired listeners in his review. In an example of one competing speaker test, he reported that a person with normal hearing could possibly gain an advantage in listening in fluctuating noise was when they can listen in between the dips in noise to gain information of speech (see masking release in Fu & Lorenzi (2006) for further reading). Masking release was not seen in listeners with mild hearing loss when a single talker competing noise was used and showed a reduction of the SRT by 7 dB SNR. Listeners with severe hearing loss could require up to 25 dB improvement in SNR to be able to score equal to their normal hearing counterparts. The same outcome was seen in other studies using fluctuating noise (Cullington & Zeng, 2008; Peters et al., 1998) and multiple-speaker babble noise (Fontan, Tardieu, Gailaird, Woisard, & Ruiz, 2015; Van Engen et al., 2014; Wilson et al., 2005).

2.5 Digit Triplet test

2.5.1 Introduction to the digit triplet test (DTT)

Traditional methods of screening for individuals aged 18 and above for hearing impairments require certified personnel and exacting procedures (ASHA, 1996). The ASHA guidelines also allow for screening for hearing disability using questionnaire-format outcome measures such as the Hearing Handicap Inventory for Elderly (HHIE) or the Self-Assessment in Communication (SAC). As both of the questionnaires are based on self-perceived deafness, one may argue that a person’s disability is subjectively different to another’s, even when absolute hearing thresholds are the same. Issues like stress level, self-motivation, stigma and emotional support could affect a person’s judgment and attitudes towards their ability to hear (Jang, Mortimer, Haley, Chisolm, & Graves, 2002; Vestergaard Knudsen, Oberg, Nielsen, Naylor, & Kramer, 2010). Therefore there is a need for an accessible and relatively economical mode of hearing screening that could provide an objective assessment of hearing sensitivity. Yueh et al. (2010) studied the effect of the design a hearing screening tool and its long term effectiveness towards motivating adults to get treatment or intervention for their hearing problems. Three types of screening strategies were investigated: tone emitting otoscopes, a hearing disability questionnaire, and a combination of both tools. In terms of the effectiveness of hearing screening as a tool for early detection, they found that adults who were screened are more inclined to go for more audiology visits and use hearing aids. As for the design of screening tools, they found participants who used a tone emitting otoscopes were more likely to go for hearing tests and use hearing aids. Using tone emitting devices alone or in combination with a hearing handicap questionnaire seemed to give higher confidence to listeners to go for hearing assessment and intervention because of its objective-like manner of testing. However in their discussion, they noted that it is possible that using hearing handicap questionnaires led to a higher rate of false positives, as the measure is less specific than the tone emitting otoscopes. It is also possible that those who were tested with the tone emitting otoscopes had higher degrees of impairment which may have led to greater need of intervention.

Currently one method of screening using speech materials is the Digit Triplet Test (Smits et al., 2004). A DTT uses three single digits (digit triplets) presented in pseudo random sequences. The test is designed to assess a person's ability to hear in noise, which is significant in identifying normal and impaired hearing listeners because most hearing impaired listeners have difficulties resolving signal and noise (Moore, 1996). Listed below are some advantages of the DTT (HearCom, 2005):

- It uses digits as stimuli which are highly familiar to native and non-native speakers of language.

- The test is presented at suprathreshold levels which allow minimum disruption from environment noise.

- The test measures SRT in noise, which is not influenced by absolute presentation level, but by relative levels of stimuli and noise over a relatively large intensity range. Therefore, no calibration is required for use in different transducers.

- It provides the user with a fast and comparatively easy assessment of his/her own auditory capability in relation to the normal hearing population.

- It allows the general practitioner to check a patient's hearing ability without having to invest into specialized equipment and training of employees to administer the test.

- It allows the provider to make a statistic about hearing ability in internet users and in specialized user groups.

- It increases public awareness of hearing impairment and treatment options for patients with a hearing problem.

The implementation of DTT has allowed researchers to produce results on the prevalence of hearing loss as well as auditory profile of test users. DTTs use digits as stimuli because they are easily recognizable, which means that minimal training is required for this test, this is why many researchers have chosen this test to be used in different languages (Jansen et al., 2010; King, 2010; Smits et al., 2004). To ensure equal intelligibility for the digits, many authors only used digits with equal numbers of syllables (Smits et al., 2004; Wagener et al., 2005). An exemption to this is the Polish

DTT (Ozimek et al., 2009) which used monosyllabic and disyllabic digits in as there are not enough digits of each type to create a homogenous list. They concluded that the inclusion of different syllables did not affect the outcome of the test. A table to compare normative values of various versions of the DTT is described in Table 42, Chapter 6 page 214.

2.5.2 Internet and telephone based hearing screening

The Dutch DTT was implemented using telephone and internet (Smits, Merkus, & Houtgast, 2006). A total of 36611 participants volunteered within the study period, with the internet-based DTT test showing almost five times more respondents compared to the telephone test over the one month period of observation. Participants who used the internet based testing were also younger than those who completed the test using telephones. Ninety-five percent of participants said the test was easy or had little difficulty performing the test. Based on a regression analysis, the chances of participants seeking help for their hearing after the test was dependent on hearing status (participants in the 'poor' group were more likely to visit professionals) and gender (male participants were less likely to follow up with the test). The initiative of testing via the internet was also well received and it managed to create increased awareness of hearing loss.

A study by Bexelius et al. (2008) on the efficiency of an internet based hearing test proposed that an internet or telephone based hearing tests cannot replace a clinical pure-tone testing by an audiologist, but is recommended to be used as an objective and cost effective screening tool for the adult population. Linssen, Anteunis, & Joore (2015) conducted a cohort study of the cost effectiveness of using different hearing screening strategies in older adults. The cost effectiveness of telephone screening, internet screening, screening with a handheld device, traditional audiometric screenings and no screening at all were compared to the cost per quality-adjusted life-year (QALY). The QALY includes costs of consultation with medical and health professionals, hearing aids and the cost of its maintenance, cost of usage of utilities (internet data, telephone bills) and annual depreciation and maintenance cost of audiological screening equipment. The quality of life was measured using a simple rating score of between 0 (death) to 1 (perfect health) using a health utility index questionnaire which was developed by the research team as a standard tool to measure hearing handicap. They found that internet screening was the most cost effective method of hearing screening compared to other methods. It was slightly more effective than telephone screening because of its lower cost and higher participation rate. Screening using a handheld devices and traditional audiometric screening were the most expensive methods of screening and were found to be less effective in improving the quality of life. Leensen & Dreschler (2013) studied real-world issues encountered when using an internet based hearing screening test. Amongst factors that could influence an internet based hearing screening test that was studied were test-retest reliability, effect of type of masking noise, binaural or monaural presentations, test environment, differences in sound processing abilities of different computers, effect of level of presentation and presence of background noise. They found that hearing screening test via the internet was reliable with a small but significant learning effect of 0.5 dB. Using low pass speech noise gave the best result in terms of consistency and separation between normal and hearing groups which showed that by unmasking high frequency sounds, normal hearing listeners could maximize the use of the information obtain from the unmasked region whereas severe hearing loss listeners would not gain any benefit at all from this release of masking. In terms of presentation mode, binaural presentations showed small advantages over monaural presentations and was consistent with previous findings in the DTT (Smits, Merkus, et al., 2006). An important question regarding the validity of testing an internet based hearing test would be whether similar performance could be obtained from listeners at home compared to lab measurements. There was a slight deterioration of results found in this study when the test, where normal hearing results were about 1 dB worse when performed at home. A possible reason for the poorer result obtained from home is the spectral differences by using different sound cards, audio settings and type of headphones. However, as the test was presented by using matched signal and noise playback, the signal-to-noise presentation should not be affected unless users attempted the test using loudspeakers in highly reverberated rooms (Culling, Zhao, & Stephens, 2005). Background noises seemed to not have any significant effect on the test as the presentation levels are at a considerable level above most common noise floor. In addition, Leensen & Dreschler (2013) recommended that monaural or binaural testing using headphones is preferred compared to loudspeakers for an internet based hearing screening test as it is more reliable especially when detrimental factors like external noise and room reverberation can be reduced.

2.5.3 Dutch DTT (Smits et al., 2004)

The first DTT as a screening tool via fixed line telephones were developed by Smits et al. (2004) in Dutch. They identified that assessing the ability of speech perception in noise was a reasonable gauge of a person's hearing impairment. Digits were used instead of words or sentences because or several reasons: (1) they are among the most frequently used words in any language; (2) digits are not easily memorized when tested in random order (3) the use of digits allows the test to be constructed as a self-administered automatic test; (4) Digits were found to produce higher slope scores compared to words which allows for greater sensitivity to the screening test. A measure of speech reception threshold in noise (SRTn) which is defined as the ability to recognize 50% of the digits correctly was used. They used an adaptive procedure introduced by Plomp & Mimpen (1979) which was aimed at improving the reliability of detecting SRTn, with the difference of adding an additional ten extra presentations compared to the original recommendation to increase accuracy. A white noise signal was used through filters to generate a speech-shaped noise feature and was used as the masker for the test.

As this was the pioneering work on the DTT as a screening tool using telephones, several questions needed to be answered which were: What are the influences of telephone use and real world listening environment in SRTn? Can the measure of SRTn be sensitive enough to be used as a standardized measure of speech perception? And finally, how does this new test compare to other established measure of speech perception in noise?

To optimize the digits, they recruited eighty normal hearing participants and tested randomized digit triplets according to participants' willingness to spend time on the tests. All digit triplets were presented adaptively and in total, 285 lists were presented. All individual scores were adjusted to the average score of all participants for every digit in each position. Only triplets with steep slopes and a specified range of SRTn were selected for the final test. An alternative method of optimizing the digits were later introduced by Melanie A. Zokoll et al. (2012) and Akeroyd et al. (2015).

To evaluate the recorded digits in different listening environments, Smits recruited ten normal hearing participants and tested them at the department using telephones, at home using participant's own telephone and at the department using headphones directly attached to the sound card. External noise influences were not measured during this experiment, but it was assumed that the listeners would use the telephone in normal listening environment and at the same time the stimuli were presented at a level higher than normal communication. They found no significant difference in using telephones at home or at the department but using headphones showed marked improvements in SRTn. The differences in score between transducers were contributed to the fact that telephones have inferior sound quality compared to headphones and more importantly it has a smaller frequency response bandwidth which is around 300 to 3400 Hz whereas headphones generally has a bandwidth between 60 to 12000 Hz. An interesting finding was that the slope scores using telephones were steeper than using headphones. Inferior sound quality could have made the task more difficult for listeners making the test more sensitive towards hearing levels. An advantage to this is that the steeper slopes using telephone has helped made the test more promising as it is able to better discriminate digit perception in listeners. The outcome of this experiment has helped to answer an important question about presenting the test over fixed line telephone, that it is possible to obtain consistent results regardless of the type of telephone handset/receiver with test-retest reliability of less than 1 dB.

Smits et al. validated the Dutch DTT by testing 76 ears with varying levels of hearing in four different conditions which are the DTT using headphones and telephones, and the speech-in-noise test using Dutch sentences using headphones and telephones. They found strong positive correlations between the Dutch DTT and sentence test in both transducers. This suggests that outcome of the Dutch DTT is consistent with established sentence test used in clinics. The SRTn of the Dutch DTT is -6.9 dB and slope of 20%/ per dB. To examine the sensitivity and specificity of the Dutch DTT, a receiver operator characteristic curve was plotted. An optimum cut-off level of -4.1dB was identified and gave the sensitivity of 79% and specificity of 100% when referred to pure tone average of 20 dB HL at three octave frequencies of 500, 1000, 2000 and 4000Hz.

As a summary this important study has shown that it was possible to develop a fully automatic screening test using digits in noise through telephones. This was only made possible as several key steps were taken which were careful recording and selecting of test materials and masking noise, optimization of audio recordings, validity of using different transducers and validation of test by testing subjects with varying hearing levels. There were several issues that must be noted. First, because of the reduced bandwidth in telephones, hearing loss that falls outside of the bandwidth may not be distinguishable in the DTT. As there were overlaps between normal and impaired hearing subjects using the -4.1 dB cut-off level, it is possible that the test could be less sensitive towards high frequency hearing thresholds. Sloping hearing losses represent a large proportion of the hearing loss in adults as noise induced hearing loss and presbycusis is dominant within this group (Abu Bakar, 2007). Secondly, the influence of extraneous noise was not fully examined. It is common to have interrupted conversations during telephone calls and also noise coming from common household items that could increase the noise levels during the test.

2.5.4 German DTT (Wagener, Eeenboom, Brand, & Kollmeier, 2005)

This version of the DTT used headphones to normalise digits in both broadband and limited band response to simulate a telephone's frequency response. Twelve normal hearing participants were recruited to normalise the digits at seven different SNRs. Normalisation improved slope of intelligibility by 5% dB with level adjustments of 2 dB and 2.8 dB for broadband and telephone use respectively. Average SRTn and slope for the German DTT for broadband headphones and telephones are -10.3 ± 0.4 dB at 17.6%/dB and -10.6 ± 0.4 at 17.2%/dB. Zokoll et al. (2012) suggested using the German DTT as a model for optimization or normalisation for future DTTs. They reviewed other established triplet tests and found the spread of standard deviation of the slope function to be smaller which produced steeper slopes. This pattern of marked improvement was found in French and British triplet tests which used the same method of normalisation. As the digits were highly optimized due to strict protocol in material selection and level adjustments, it is possible to separate language and speaker specific influences in the DTT and allow for cross language comparisons of the test. As their recommendation, DTT normalisation should be done on each digit individually at each digit position (front, middle and back) especially for languages with large differences in prosody for digit pronunciation at different digit positions. A comparison across DTT versions in different languages was also done in this study.

2.5.5 Polish DTT (Ozimek, Kutzner, Sęk, et al., 2009)

As mentioned earlier, the digits selected for the Polish DTT were not homogenous as half of the Polish digits are disyllabic. They produced four triplet lists containing 25 unique triplets and special consideration were given to ensure equal SRT, slope scores and phonemic balance for each list. A significant difference between the masker used in this test compared to the Dutch test was that the digits itself were used to generate the masking noise by superimposing all recorded triplets. The lists were designed to be statistically equivalent with a standard deviation between lists of not more than 1 dB. Participants were required to key in the digits twice to avoid mistakes and/or poor concentration. List equivalency was tested by recruiting twenty normal hearing listeners. Analysis of variance showed no significant list effect in the Polish DTT suggesting that the lists were equally intelligible and could be used as comparable sets. The average slope scores for the lists in the Polish DTT were found to be close to the German test and higher compared to the Dutch DTT. The differences were attributed to the type of noise used as a masker for the Polish DTT. Average SRT for the Polish DTT using headphones was -9.4 dB with an average slope of 19.7%/dB.

2.5.6 French DTT (Jansen et al., 2010)

The French DTT was developed, evaluated and validated for broadband headphones and telephone use. In the process of selecting the suitable digits to be optimized, they also studied the correlation between results obtained from headphone and telephone use. They found a low statistical correlation between the two groups which carries a strong argument to perform optimization separately for both transducers. The average SRTn for headphone and telephone were -10.5 dB and -6.4 dB respectively, whereas the slope values were 27.1%/dB and 17.1%/dB. The methods used to normalise the digits were similar to the methods introduced in the German DTT. This resulted in the steep slopes which is actually the highest compared to other versions of the DTT listed in this review. As for the implementation of the pass and refer outcome, they have chosen to divide the SRT results into three categories which are good, insufficient and poor similar to the changes made in Smits & Houtgast, (2005). This is due to the large overlap between normal and hearing impaired listeners. This overlap is mostly contributed by good hearing at low frequencies followed by mild level of hearing loss of less than 40 dB HL at high frequencies. This pattern is evident in all the digit triplet tests using a test specific noise listed in this review. The authors advertised the test and within the first month of its introduction over 18000 completed calls were recorded. They excluded data that were not suitable (such as when the caller age was outside the study range), leaving about 15391 calls that were analysed. The data allowed them to set up profiles and help seeking behaviour of self-motivated callers based on demographics and outcome of the screening test. The same group of researchers conducted a follow up study to compare the French DTT to an established everyday sentence test in French and the newly designed matrix sentence test in French (Jansen et al., 2012). They found significantly strong positive correlation between the tests, supporting the reliability of the tests as well as the consistency of the DTT as a screening tool in identifying hearing impairment in noise.

2.5.7 American English DTT/US National Hearing Test (NHT) (Watson, Kidd, Miller, Smits, & Humes, 2013; Williams-Sanchez et al., 2014)

Following the huge reception received in several countries in Europe, a group of researchers in the United States of America (USA) attempted the implementation of DTT as a standardized national screening tool. Watson et al. (2012) developed and evaluated the DTT and Williams-Sanchez et al. (2014) validated the test by studying SRTn amongst veterans through the Veteran Affairs office. Watson et al. (2012) used similar method of optimization that was used in the Dutch DTT which was to identify triplet as a function instead of measure each digit at each position of the triplet. Ten normal hearing listeners were recruited for this process to optimize 160 triplets. After the process of eliminating inappropriate triplets on the basis of slope uniformity and steep psychometric functions, 64 triplets were found to be functional for the test with an average SRTn of -7.4 dB triplet score. Using randomized sequence of the triplets, they have also found that testing 15 triplets was adequate to produce a reliable measure of SRTn. Using -5.7 dB as a cut-off level for refer criteria, the sensitivity and specificity of the NHT was 80 and 83 percent respectively.

Williams-Sanchez et al. (2014) validated NHT by recruiting more than 500 veterans with varying levels of hearing. The NHT was also compared to other speech-in-noise tests. First, the NHT was found to have stronger correlation to the average hearing thresholds of 4 frequencies (500, 1000, 2000 and 4000 Hz) than to 3 frequency averages (500, 1000 and 2000 Hz). The statistical correlation between the SRTn and the 4 frequency average hearing thresholds were comparable to the Dutch and French versions of DTT. Secondly, although the sensitivity of the test was almost equal to that measured by Watson et al. (2012), the specificity in this study dropped. This could be contributed by the low number of normal hearing volunteers in this experiment. Despite this, they reported the NHT was comparable to other speech tests as well as other established DTT versions and was quick and easy to administer.

2.5.8 Australian English DTT (Golding, Seymour, Dillon, & Carter, 2007)

The first Australian English version of the DTT (Telscreen version I) was launched in 2007 and used the Dutch DTT test a model for development (Golding et al., 2007). When compared to the four frequency average of 75 participants with various levels of hearing, they found a significant but moderate positive linear relationship. A second version (Telscreen version II) was implemented to see if the test could be improved. They used a speech shaped noise that was modified both spectrally and temporally. Version II showed higher correlation (r = 0.77) to hearing thresholds compared to version I (r = 0.63). This suggests superior outcomes using version II using the in-house developed masker.

In one study, Telscreen II was used to investigate help seeking behaviour amongst callers of the test (Meyer et al., 2011). Participants were interviewed on several aspects of issues related to hearing loss and hearing aids. They found that out of 193 participants who attempted and failed the Telscreen II test, only 36% sought help for their possible hearing loss, with the majority of those opting to visit hearing service providers such as audiologists and hearing aids dispensers. Less than half of the 36% of people who sought help were recommended hearing aids with only 9% of respondents actually deciding to get hearing aids. A regression analysis showed participants who had thought of the possibility of wearing hearing aids were found to be more likely to get hearing aids in this study. It is difficult to measure if the Telscreen II test actually motivated participants to seek help and if the test had any influence in decision making as other variables could affect this. It is very clear though that hearing aid acceptance is very low within the group who failed hearing screening. The authors noted that more research needs to be done to understand if the screening test is a cost effective tool to detect hearing loss and ultimately lead to audiological intervention as the success rate of hearing acceptance was small in this study. They also noted future research should also include other internal and external elements that influence decision making to better understand patterns that could motivate or discourage hearing impaired patients into accepting hearing aids.

2.5.9 New Zealand Hearing Screening Test (King, 2010; O'Beirne, King, Murray, Fromont, & King, 2012)

Work on developing the DTT in New Zealand started in 2010 by postgraduate students at the University of Canterbury in Christchurch. This research group, led by Assoc. Prof. Greg O'Beirne, has developed and normalised the New Zealand English DTT, and also the Māori version of the DTT (Murray, 2012; O'Beirne et al., 2012). As they are produced in two of the official languages in New Zealand, these two tests are referred to as the New Zealand Hearing Screening Test (NZHST).

The English version was developed and optimized for New Zealand English speakers (King, 2010). They applied the German DTT model of digit selection and optimization but test was delivered using custom software written by O'Beirne. The mean SRTn for normal hearing participants was -12.8 dB SNR ± 0.4 with an average slope of 17.3% dB, which was similar to the German test. The monaural evaluation of the SRTn in the NZHST showed significantly strong positive correlation to average pure tone thresholds (r = 0.809). Using the calculated receiver-operator curve (ROC), the chosen cut-off value of the English test was -10.4 dB SNR, which yielded a test sensitivity of 88% and specificity of 81%.

The Māori version was developed in part to improve access to hearing screening among Māori and to obtain more information regarding their hearing status (Murray, 2012), as previous census study showed a higher prevalence of hearing lossamong this population. Disyllabic Māori digits were used in this test (the monosyllabic digit 4 [wha] was omitted) and the normalisation procedure was conducted at four different SNRs by recruiting 8 fluent speakers of te reo Māori. The average level adjustment for this test was 0.7 dB, which produced a predicted SRTn value of -11.3 dB ± 1.0 dB and a predicted slope of 15.77%/dB. The standard deviation after normalisation in this test is high compared to other version of the DTT.

This could have been contributed to by possible lapses in concentration and errors in the responses in the small number of participants. Due to time constraints and the limited number of participants, an evaluation of the triplet lists in the Māori version could not be done, and so as of 2016 the test is not yet available for public use.

Bowden (2013) reviewed the NZHST and attempted to improve the homogeneity of test lists in the English version by ensuring equal distribution of digits within each list and at the same time evaluate the Māori version. Previously in the first edition of the English test, higher numbers of digits with steeper slopes of intelligibility were included in each list to increase the slope of the test overall, but this raised the possibility of the unequal digit distributions being noticed by listeners and biasing their results. Correlation between the SRTn and pure tone hearing thresholds of the new version of the test was found to be similar to the previous version ($r = 0.81$). Using ROC analysis, and the cut-off for normal hearing at -10.0 dB, the test yielded a test sensitivity of 90% and specificity of 88%. Due to limited number of hearing impaired participants for the Māori test ($n = 4$), only a pilot evaluation study was conducted, which yielded a significant but moderate positive linear correlation ($r = 0.63$) between SRTn and binaural average pure tone hearing levels.

Bowden (2013) also included a spectral and temporal gap (STG noise) masking noise to investigate the possible improvement of SRTn in normal hearing listeners in the English version. As expected, using STG improved the results in normal hearing participants as normalisation procedures produced expected average SRTn of -11.5 dB SNR compared to -8.9 dB SNR in steady state speech shaped noise. This study suggested the potential of using the STG noise as a way to improve sensitivity and specificity of the test. More information is needed to investigate the influence of this noise within groups of listeners with varying levels of hearing.

As part of this continuing research at the University of Canterbury, this study aims to develop, optimize and evaluate a Malay version of the DTT for telephone and internet applications using headphones. The Malay version will use a test specific noise as well as the STG noise.

2.5.10 Updates on DTT

The Dutch DTT set a new standard in hearing screening where thousands of listeners were able to screen their hearing reliably. In 2005, Cas Smits and Tammo Houtgast described the outcome of the nationwide implementation of the Dutch DTT a year after it was introduced to the public (Smits & Houtgast, 2005). The analysis was done for all callers from January to April in that year. As the test was open to the public, a substantial number of callers used mobile phones to conduct the test. The results obtained from user using mobile phones gave significantly poorer results than landlines. This prompted the author to exclude the data obtained from mobile phones and other unknown telephone types. Out of the total 65 924 callers, only 39 968 were usable in this study. It was hypothesized that the data collected from mobile phones could most likely be corrupted by external noise and poorer sound quality as users may not be indoors and could have poor reception that could influence the outcome of the test. The results showed most callers were aged more than 44 years old with a median age of 56 and 54 for men and women, respectively. Consistent with other finding of hearing loss increasing with age (Abu Bakar, 2007; Degeest, Keppler, & Corthals, 2015; Rönnberg, Hygge, Keidser, & Rudner, 2014), the study showed increasing SRTn with age both in male and female callers. They also compared the SRTn to caller's subjective perception of their hearing loss and found older callers tend to overestimate their hearing ability in noise and also the higher the age group and SRTn levels, the more variance in responses were obtained. The reliability of the test was good as the measurement error was still within 1 dB when the cut-off limit was set -4.1 dB for the good or pass criteria. A prospective study using DTT in Dutch (Stam et al., 2015) showed reduced speech perception abilities on repeating the DTT after 5 years. Participants between 18 and 70 years of age had deteriorations in their SRTn and drastic changes that were most evident in subjects aged fifty years and older. Using DTT and a questionnaire, this study was able to discuss the relationship between level of education, income and speech perception abilities in noise where no significant relationships were found. This was done due to the initial finding by Stam, Kostense, Festen, & Kramer (2013) that participants with lower education levels and income were more likely to have a hearing impairment. The DTT in Dutch was also used by Pronk et al. (2013), where an average decline of 0.18 dB per annum was found in participants aged 57 to 93 years. This value is similar to the deterioration of

0.82 dB SRTn found by Stam et al. (2015) over a period of five years (i.e. 0.16 dB per annum) for a similarly aged group of subjects. Although these differences in SRTn are small, the slope of the intelligibility function means they have a significant impact on the listeners (Stam et al., 2015) and further investigation needs to be done to explore the relationship between deteriorations of SRTn and hearing disabilities. A reasonable explanation is that the deterioration in SRTn is due to the decline of information processing speed due to age as more memory is dedicated towards perception and less from information storage (Pichora-Fuller, 2003).

2.5.10.1 Comparing optimization methods used to develop the DTT

After discovering the standard deviation of SRTn increased with hearing loss in the Dutch DTT, Smits & Houtgast (2006) attempted to investigate the variations in slope, guess rate and other possible factors that could influence the reliability of the DTT. The aim was to find possible ways to improve the adaptive measurement so that it could be used in other comparable speech-in-noise tests. By looking at more recent and numerous data collected since its implementation, they claim that it was possible to refine the level adjustments to improve homogeneity of the triplets. Here they defined homogeneity as equality of the SNR at the target point (which in the case of DTT is at 50% score) rather than the equality of steepness of the triplets intelligibility function. The refined level adjustments resulted in only a slight improvement of the standard deviation of the SRTn, which implied that the increase of standard deviation in SRTn with increasing hearing loss was not contributed by any heterogeneity of the triplets. Using mathematical simulations, they investigated the possibility of optimizing triplets by selecting digits of equal steepness within the triplet and observed possible improvements. This led to an average reduction of estimated standard deviation of the SRTn by up to 1 dB. This makes a strong argument for using the methods of optimization used in the German DTT which applied this method. As mentioned earlier in the review of the German DTT, a structured optimization procedure was suggested by Zokoll et al. (2012). Compared to the method used in the Dutch and American versions of the DTT, this approach seem to work best when the DTT was designed to use different transducers and masker as a minimal number of trials can be evaluated as the digits are measured individually during this process. It also gives the more flexibility and control over the choice of digits to be used in the triplet lists as well as high accuracy in the final triplet scores. The drawback to this method is the development and evaluation will take longer to complete as the various stages have to be completed and more participants are required to optimize and validate the test. The optimization method used in the Dutch and American DTT relies on the construct of pre-set triplets and the process of eliminating triplet outliers and finding uniformity within the triplets after testing normal hearing participants. This is a more direct approach in normalizing the triplet and possibly takes less time to construct. One disadvantage in this approach is that the final intelligibility function for participants may vary from one test to the other as no control was made towards sequence of presentation and so as a result, the SRTn and slope score could show slightly higher variation compared to the approach used in the German DTT (Zokoll et al., 2012).

2.5.10.2 Improving the sensitivity of DTT in sloping sensorineural hearing loss

The DTT was successfully implemented in several countries, as a result more information was gathered which provided additional information such as the profile of help seekers and the relationship between hearing and work (Nachtegaal et al., 2009). It has allowed researchers to understand better the clientele of the test as well as possible ways of improving the accuracy and reliability of the test (Smits & Houtgast, 2006, 2007). In this part of the review, we discuss the major weakness of the test, which is its less-than-optimal ability to identify listeners with sloping hearing loss. It is now well documented that the established DTT is very sensitive to hearing levels, however a single cut-off level for "pass" and "refer" was found to be inadequate to optimally separate between normal and impaired hearing groups especially those with sloping hearing losses. Hence, all authors of the tests agreed to include a third category which is now referred to as "insufficient". This category usually refers to the range of the test where there is an overlap between normal and impaired hearing. Callers under this category will be advised to seek professional help regarding their hearing. To enhance the ability of normal hearing listeners in identifying triplets in noise, Smits & Houtgast (2007) observed SRTn changes using the Dutch digits in four types of noises which were: (1) continuous speech-shaped noise; (2) "16 Hz interrupted noise" by modulating the continuous speech-shaped noise by a 16 Hz square wave; (3) "32 Hz interrupted noise" by modulating the continuous speech-shaped noise by a 32 Hz square wave; and (4) “speech following noise” which was constructed by adding noise burst to each digit. Noise (2) showed the lowest recorded SRTn compared to the other three noises in normal hearing subjects. The result was expected as this was due to masking release in normal hearing listeners. The spread between normal hearing and hearing impaired listeners was also the highest in noise (2) and the lowest in noise (1). Analysis using receiver-operator curves in noise (2) showed improved sensitivity and specificity of the digit SRTn test in identifying hearing impaired listeners. As noise (2) was developed and tested using single digits, a more structured study was needed to see the influence of the 16 Hz masker noise in SRTn using digit triplets.

A review of the relationship between SRTn and hearing threshold levels across all DTT literature shows that hearing impaired callers within the ‘insufficient’ group display good hearing at low and mid frequencies and mild hearing loss at high frequency regions. This suggests that DTT is less sensitive towards high frequency hearing thresholds, which could lead to high false negative figures. The internet version of the Dutch DTT was compared to two other internet-based hearing screening tools called the Earcheck (EC) and the occupational Earcheck (OE) (Leensen, de Laat, & Dreschler, 2011) with participants with normal hearing and sloping hearing losses. The DTT showed a reduced sensitivity of 55%, which was considerably lower than that reported in the initial findings (Smits et al., 2004). This would suggest that even by using a broadband headphone, the DTT was still poor at detecting the highly sloping hearing losses which are common in noise induced hearing loss patients.

In this study a spectral and temporal gap noise was developed and normalised to investigate possible spread of SRTn between normal and hearing impaired groups. The noise is specifically designed to allow release from masking for normal hearing listeners hence improve the specificity of the test.

2.5.11 Summary of DTT review

The implementation of DTT as a screening tool has been successful in acheiving its aim, which is to provide an objective self-administered hearing screening test that is efficient and easy to perform. Implementing such a test does not only help callers to detect possible hearing loss but also creates awareness regarding hearing impairment, especially when supplementary information can be gathered whilst conducting the test especially in the internet version (Royal National Institute for Deaf People, 2015).

The test is especially accurate due to careful selection of digits, controlled testing and normalisation of those digits, and presentation of the test in signal-to-noise mode where the issues of standardization or calibration are largely eliminated. There are at least two methods of normalisation used in previous studies, with the approach taken by the German DTT offering better flexibility and accuracy after normalization. Performance in DTT is highly dependent on low frequency hearing which leads to a weakness in evaluating sloping hearing loss, especially when listeners have good low frequency hearing and mild hearing loss in high frequency regions. Several researchers have attempted to improve test sensitivity by improving digit selection and/or by using a different type of masking noise to encourage release from masking in normal hearing participants, with most of them observing improvements in SRTn and test sensitivity.

Due to its great potential as an assessment tool for speech perception abilities in noise, Smits, Goverts, & Festen (2013) proposed the use of digit triplets as a diagnostic tool in clinics in Dutch. The study was used in combination with data obtained from the Dutch National hearing screening test to identify the relationship between the newly proposed diagnostic tool and the well established DTT screening test. The result if this study showed that only one training list was necessary to overcome training effects and that the test could be performed in about 3 to 4 minutes. The application of digit triplets as a diagnostic tool was an interesting proposition as it allows for a broader target listener group – from normal to severe-profound hearing loss and from native to non native speaker of the language.

Currently the DTT is not recommended for use over mobile phone network because there are higher risks of corrupting factors that could affect the result such as use of the test outdoors (possible increase in noise) and poor network reception. To overcome this, current smartphones are able to use downloaded or web-based apps as an alternative. This eliminates the possibility of poor network connections affecting performance, and by ensuring headphones are connected during the test external noises can also be reduced. Additionally, the microphone in the phone can be used to monitor ambient noise. Upon permission of users, the application can assist in providing more information regarding hearing care as well as locating nearby hearing professionals using global positioning system (GPS) or mobile network triangulation.

2.6 Matrix sentence test

2.6.1 Introduction to the Matrix sentence test (MST)

Hearing aid fitting is an integral part of an audiologist practice. Recent developments in sound processing have made fitting hearing aids a more complex process where many variables in the hearing aid can be programmed and adjusted according to the patient's needs. This often requires several visits either to address certain issues regarding hearing aid use, verification of gain or validation of overall fitting. This indicates that a large amount of speech material is necessary to avoid repetition or any learning effects. This issue could be overcome by using a matrix sentence test, first designed by Hagerman (1982) in Swedish. A matrix sentence test is a semantically unpredictable sentence in a fixed grammatical structure. A selection of words belonging to each syntactic category in a sentence can be used interchangeably during each test, creating a speech test that is nearly impossible for a patient to memorize. The typical matrix sentence arrangement in most languages is [Noun + Verb + Number + Adjective + Object] but in the Spanish and French versions of this test, the Object noun precede the adjectives (Hochmuth et al., 2012; Jansen et al., 2012). There are other speech in noise tests available in other languages, such as the English, Mandarin and Malay HINT sentence tests (Nilsson et al., 1994; Quar et al., 2008; Wong et al., 2008), the Dutch speech reception test (Plomp & Mimpen, 1979), or the American SPIN sentences (Kalikow et al., 1977) which are presented either as meaningful daily sentences or unpredictable sentences (SPIN). The MST has the advantage (as does the DTT) that it can be presented as an open or closed set test where the words used are limited and each test list usually includes all the speech materials available. The matrix sentence test is designed by applying the principles psychometric theory, which should allow clear distinction of normal and impaired hearing listeners. Development of the test involves normalizing the audio stimuli to ensure all the speech stimuli used have equal difficulty, and therefore results are repeatable andare ultimately dependent on listeners ability to discriminate speech in noise.. All matrix tests use adaptive algorithms to avoid floor and ceiling effects as well as to calculate the SRT more efficiently. This allows clinician to evaluate and compare the MST test results of the same listener to observe changes in their speech perception. As the development of the matrix sentence test is fairly well standardized, comparison between versions of this test in multiple languages is possible. Zokoll et al. (2013) evaluated the matrix sentence tests in Russian, Turkish, Spanish, British English, Danish, Dutch, French, German, Polish and Swedish. As the tests are methodologically similar, they found that the SRTs in noise were comparable between the tests with an average of -8.2 dB and a standard deviation of 1.28 dB. They found slightly larger variation in slope of the SRT with an average slope of 14.5 %/dB and a standard deviation of 3.26 %/dB, which could be due to language and speaker specific dependent factors in speech intelligibility.

2.6.2 Development protocol for the matrix sentence test

Acknowledging the importance of standardizing this popular test, the International Collegium of Rehabilitative Audiology (ICRA) has recently published a guideline to ensure new versions of this test will be comparable to the ones currently available (Akeroyd et al., 2015). A summary of the recommended guidelines is shown in the table below:

Table 3: ICRA recommendations in the development of an internationally comaparable matrix sentence test

Stages of development	Recommendations by Akeroyd et al., (2015)
General construction	1. To include base matrix of 50 words in a semantically fixed sentence structure that is appropriate to the language.
Word selection	1. Homogenous number of syllables for each word. 2. Representative of language specific phoneme distribution. 3. Semantically neutral words.
Speaker	1. Native speaker and not necessarily formally trained. 2. Natural intonation 3. Use female speaker if possible as it is a compromise between an adult male and children's voice.
Recording & editing	1. Record 100 sentences to account for co-articulation by securing word pairs of all possible word combination. 2. Ensure each word occurs equally across all generated

		sentences.
Masking noise	1.	A stationary or steady state noise that contains the same long term spectrum as the speech materials.
Normalisation	1.	Recruit at least 10 normal hearing native speaker of the language.
	2.	Measure speech intelligibility at fixed SNRs that covers 10% to 90% at speech level (e.g. 65 dB SPL).
	3.	Ensure training lists of at least 40 sentences for each subject prior to testing.
	4.	Adjust word levels to reach mean SRT as target at a limit of ± 2 to 4 dB.
	5.	Separate measurement for specific tests such as telephone use, processed speech materials and noise type.
Evaluation	1.	Test list equivalence.
	2.	Obtain normative data.
Validation	1.	A multi-centre study with normal and hearing impaired listeners.

In a review of multiple languages produced under the Hearcom initiative (hearcom.com.eu), Kollmeier et al. (2015) described their experience in the development of the MST. The review provided recommendations similar to those produced by ICRA, however a strong emphasis was given to explaining differences between the tests that they have developed and providing recommendations for future developments of other MSTs. Using the ICRA framework shown above, the recommendations by Kollmeier et al. (2015) are listed in the table below:

Table 4: Specific recommendations for the development of the matrix sentence test by Kollmeier et al. (2015)

Stages of development	Recommendations by Kollmeier et al. (2015)
General construction	1. When possible, describe phoneme distribution of the selected words in comparison to the corpus of the respective language.
Word selection	2. Consider using the same grammatical rules that involve the variation of the form of a noun, pronoun, verbs and adjectives which have gender dependence. For example, use the same tense in all sentences as a change in tense can alter gender references between sentences for languages with gender dependencies.
Speaker	1. No specific recommendations
Recording & editing	1. Speaker should exhibit an idealized or official pronunciation of the language. Average rate of about 4.1 syllables/second was found across nine MSTs. 2. Train speakers to keep speech effort constant during recording. 3. 'In' and 'out' edit cuts should start at 0 degrees and end at 180 degrees of phase. Out phase cuts were made by including spectral contents of successive words. 4. To make speech sounds more natural a fixed overlap between successive words are recommended (e.g. 15 milliseconds) 5. Two cutting approaches are possible; i. Cut all recordings into five single words, or ii. Cut all recordings into single words except the last two words in the sentence top allow natural sound for speech parts that do not require mixing.
Masking noise	1. Use a quasi-steady state speech-shaped noise by superimposing speech materials for at least 30 times. 2. When possible, compare the speech-shaped noise to the long term average speech spectrum (LTASS) (Byrne et al., 1994).
Normalisation or optimization	1. Use the following logistical function for word specific intelligibility function, $SI_{word}(SNR) = 1/(1+exp(4S_{50word}(SRT_{word} - SNR))$ With, SRT_{word}: word-specific SRT in dB at 50% intelligibility S_{50word}: *slope* at SRT_{word} in %/dB

	2. Report level adjustments which can be varied across languages (Maximum adjustment for the nine MST tests in Hearcom project is ± 3 dB).
Evaluation	1. Evaluation should be done separately for open and closed set type response
	2. Measure training effect using adaptive procedure (Brand & Kollmeier, 2002) measuring 50% intelligibility with lists of at least 20 sentences.
	3. SRTn and slope calculated for each list by measuring responses at 2 or 3 fixed SNRs in about 20 to 80% intelligibility.
Validation	1. Measure test-retest reliability in two separate sessions.
	2. Use adaptive measure to record SRTn in normal and impaired hearing listeners.

The review and comparison by Kollmeier et al. (2015) of various MST versions will be described in 2.6.4.

2.6.3 Summary of features and reference values of various MSTs

The multilanguage matrix sentence tests were developed using similar methods. Most of the work is from (or at least co-authored by) the group at Hearcom, and is in European languages. To assist readers, the MSTs are listed alphabetically and are presented in point form as there are many similarities between tests and also some distinct features in each test that are highlighted here. A table to summarize and compare different normative values between different versions of the MST are shown in Table 47, Chapter 6, page 230.

2.6.3.1 Danish MST/Dantale II (Wagener et al., 2003)

i. Introduced speech material editing methods that preserved co-articulation.
ii. Female speaker.
iii. Sentence structure: Name + Verb +Numeral + Adjective + Object
iv. 16 lists of 10 sentences that could be combined to 10 lists of 20 sentences.
v. SRTn: fixed SNR measurement, -8.4 ± 0.2 dB (across lists).
vi. Slope: 13.2 ± 0.8 %/dB (across lists).

vii. Learning effect was 2.2 dB between list 1 and list 2 but reduced to less than 1 dB after two training lists of 20 sentences.

2.6.3.2 Dutch MST (Houben et al., 2014)

i. Female speaker.
ii. Sentence structure: Name + Verb +Numeral + Adjective + Object
iii. 14 lists of 20 sentences.
iv. SRTn: fixed SNR measurement, -8.4 ± 0.2 dB (across lists; ± 0.7 dB across listeners).
v. Slope: 10.2 ± 0.9 %/dB (across lists); -10.5 ± 1.4 %/dB (across listeners).
vi. To evaluate the lists, 45 participants were recruited from three different centres with different headphones and sound cards. As no significant difference was found between centres, this test can be used in the Netherlands and in Flanders.

2.6.3.3 French MST (Jansen et al., 2012)

i. Female speaker.
ii. Sentence structure: Name + Verb + Numeral + Object + Colour
iii. SRTn: fixed SNR measurement, -6.0 ± 0.1 dB (across lists; 0.6 across listeners).
iv. Slope: 14.0 ± 1.6 %/dB (across lists).
v. Learning effect using word and sentence scoring was evaluated. Word scoring learning effect was ± 1 dB from first to second lists and 0.4 dB between list 2 and 3.
vi. Significantly strong positive linear correlation between French MST SRTn and pure tone averages of 250, 500, 1000, 2000 and 4000 Hz in participants with varying hearing levels (r = 0.94).

2.6.3.4 German MST/Oldenburg Satztests (OLSA) (Wagener, Brand, & Kollmeier, 1999; Wagener, Kuhnel, & Kollmeier, 1999; Wagener, Kuhnel, Kollmeier, et al., 1999)

i. Published in German
ii. Male speaker.
iii. 12 lists of 10 sentences.

iv. Sentence structure: Name + Verb + Numeral + Adjective + Object
v. Intelligibility increase by an average of 3%/dB.
vi. SRTn: Fixed SNR measurement, -7.1 ± 0.16 dB (across lists; 1.1 across listeners). Adaptive measurement, -6.8 dB (open and closed set)
vii. Slope: 17.1 %/dB.
viii. Learning effect was between 1 to 2 dB. Suggested two training lists before measurement of SRTn.

2.6.3.5 Italian MST (Puglisi et al., 2015)

i. Female speaker.
ii. Evaluated both in open and closed set response.
iii. 12 base list of 10 sentences or 6 lists of 20 sentences.
iv. Sentence structure: Name + Verb +Numeral + Object + Adjective
v. SRTn: fixed SNR measurement, -7.3 ± 0.2 dB (across lists); -7.4 ± 0.9 dB (across listeners).
vi. Slope: 13.3 ± 1.2 %/dB (across lists); -14.3 ± 3.6 %/dB (across listeners).
vii. Learning effect for open set responses between presented list 1 and list 2 was 1.2 dB and decreased to 0.5 dB from list 2 to list 3. For the closed set responses, effect was observed from list 1 to list 2 by 1.1 dB and reduced to 0.3 dB from list 2 to list 3.
viii. Test-retest reliability of 0.5 dB for open set and 0.6 dB for closed set responses.

2.6.3.6 Polish MST (Ozimek, Warzybok, & Kutzner, 2010)

i. Male speaker.
ii. Sentence structure: Name + Verb +Numeral + Adjective + Object
iii. Words chosen from top 500 most frequent words used in the language.
iv. Evaluated both in open and closed set response.
v. SRTn: fixed SNR measurement, -9.6 ± 0.2 dB (across lists); -8.0 ± 0.4 dB (sentence scoring); adaptive measurement, -8.0 ± 1.3 dB (open and closed-set).
vi. Slope: 17.1 ± 1.6 %/dB (across lists); -21.8 ± 2.8 %/dB (sentence scoring).
vii. Comparison between test with and without experimenter using an adaptive 1-up/1-down staircase procedure showed no significant differences in SRTn.

2.6.3.7 Russian MST (Warzybok, Zokoll, et al., 2015)

i. Female speaker.
ii. Sentence structure: Name + Verb +Numeral + Adjective + Object
iii. Evaluated both in open and closed set response.
iv. SRTn: fixed SNR measurement, -9.5 ± 0.2 dB (across lists); -9.4 ± 0.7 dB (across listeners); adaptive measurement, -8.8 ± 0.8 dB (open-set); -9.4 ± 0.8 dB (closed-set).
v. Slope: 13.8 ± 1.6 %/dB (across lists); 14.0 ± 3.4 %/dB (across listeners).
vi. Significant learning effect observed in both open and closed-set responses between lists 1 & 2 compared to other lists. This suggests training of two lists prior to test evaluation.
vii. Using presentation noise level of 80 dB SPL resulted in decreased speech intelligibility compared to 45, 55 and 65 dB SPL. Author suggests presentation level of MST should not exceed 75 dB SPL.

2.6.3.8 Swedish MST (Hagerman, 1982)

i. Female speaker
ii. Sentence structure: Name + Verb +Numeral + Adjective + Object
iii. Words were recorded and edited individually producing synthesized word by word pronounciation. A new approach to preserve speech transitions between words was later introduced in the Danish MST (Wagener et al., 2003).
iv. Word optimization level adjustment not more than ± 1.3 dB.
v. SRTn: fixed SNR measurement, -8.1 ± 0.3 dB (across lists); -8.2 ± 0.7 dB (across listeners).
vi. Slope: 16.0 ± 3.4 %/dB (across lists).
vii. Learning effect was studied by comparing intelligibility scores (%) from lists 1 to 5. Significant learning effect found between lists 1 and 2.

2.6.3.9 Spanish MST (Hochmuth et al., 2012)

i. Female speaker
ii. Sentence structure: Name + Verb + Numeral + Object + Adjective (only male objects used due to gender dependent declension of adjectives in Spanish)

iii. SRTn: fixed SNR measurement, -6.8 ± 0.13 dB (across lists); -7.7 dB (closed-set); adaptive measurement, -6.2 ± 0.8 dB (open-set); -7.2 ± 0.7 dB (closed-set).
iv. Slope: 13.2 %/dB (open-set); 14.0 %/dB (closed-set);
v. Learning effect was significant for open and closed-set tests at 1.1 dB and 0.8 dB respectively from first to second lists and 0.5 dB between list 2 and 3.
vi. No significant difference was found between Spanish speaking subjects from different centres.

2.6.3.10 Turkish MST (Zokoll et al., 2015)

i. Female speaker.
ii. Sentence structure: Name + Numeral +Adjective + Object + Verb
iii. SRTn: fixed SNR measurement, -8.3 ± 0.18 dB (across lists); adaptive measurement, -7.2 ± 0.8 dB (open-set); -7.9 ± 0.8 dB (closed-set).
iv. Slope: 14.1 ± 1.0%/dB (across lists).
v. 12 lists of 10 sentences or 6 lists of 20 sentences.
vi. Learning effect was significant for open and closed-set tests at 2.0 dB and 1.4 dB respectively from first to second lists and 0.5 dB between list 2 and 3.
vii. Mean SRTn for the open-set response format in quiet was 20.3 ± 4.1 dB when measured adaptively and correlated more closely to pure-tone audiometric thresholds rather than speech reception thresholds in noise.

2.6.4 Updates on MST

The various MST versions showed that the test has high accuracy and reliability. The standard deviation in measurement of SRTn for all the MST across test lists were less than ± 0.5 dB and test retest reproducibility in clinical settings of about ± 1.0 dB in normal hearing participants. This is possible as attempts were made to ensure test increased in its sensitivity by the process of optimization and ensuring measurement were least influenced by participants' ability to learn the materials used. From the information above, optimization has improved test sensitivity by as small as 3%/dB in the German MST (Wagener, Kuhnel, Kollmeier, et al., 1999) and as high as 8%/dB in the Spanish MST (Hochmuth et al., 2012). The inconsistency of improvements after normalisation could be attributed to differences in language used, technique of generating masking noise and unique speaker differences (e.g. individual voice traits) (Ozimek et al., 2010).

The relationship between the SRTn at a fixed noise level of 65 dB SPL and pure tone audiometric thresholds were analytically examined by Wardenga et al. (2015). They also attempted to calculate the level of hearing threshold at which noise levels have no effect on the SRTn. They examined the relationship between SRTn using the Oldenburger Satztest (OLSA) in 315 ears with hearing levels ranging from -5 to 90 dB HL by systematically grouping listeners into five groups. They observed significant but moderate linear relationships between SRTn in OLSA within corresponding groups. Using separate linear regression analysis for the groups, an intersection of the linear domains was found at a pure tone audiometry threshold of 47.6 dB HL. They concluded although it was acceptable to conduct MST at a single noise level for all listeners, it was more appropriate to perform MST at 65 dB SPL noise level only for listeners with pure tone average of 47 dB HL and lower at inter octave frequencies of 500 Hz to 3000 Hz. Listeners with hearing levels exceeding 47 dB HL at frequencies of 500, 1000, 2000 and 3000 Hz were recommended to be tested using the MST in quiet as no difference of testing with or without noise was observed in this group. As described earlier, due to the language and talker-specific differences of the MST across languages, this cut-off level is most likely applicable to OLSA and native speakers of German only, while other versions of MST would require a specific investigation for a normative reference.

One particular issue that has been reported in all of the MSTs above is training effects. This is due to the multitasking nature of this test whereby the listener has to listen to speech stimuli in noise, choose words that are heard from a combination of possible words and respond appropriately. If listeners can familiarize the response type and test materials used, it is likely to produce a more consistent result and at the same time improve test sensitivity. In the MST studies mentioned above, the SRT reduced significantly with SRTn scores difference between list 1 and 2 of up to ± 1.2 dB. Training effect using open-set responses were higher (up to 2 dB) compared to closed-set responses however differences in training effect between open and closed set also reduces almost equally after completion of the second list (Hochmuth et al., 2012; Puglisi et al., 2015; Warzybok, Zokoll, et al., 2015). As a summary, authors recommended that listeners should complete at least 2 lists of 20 sentences each before a reliable measure of SRTn can be made in any MST evaluations, either closed-set or open set.

Comparison between different versions has helped researchers understand the test better. In terms of the differences in SRTn and slope, Kollmeier et al. (2015) reviewed currently available MSTs within the Hearcom project. They noted the spread of 4.1 dB between the highest and the lowest SRTn found in various MSTs. They observed that Slavic languages (Polish and Russian) tend to show lower SRTn compared to Romanic languages (Spanish or French). Since the approaches to developing all of the tests were similar, they contributed the difference in the SRTn due to language and speaker differences. This is because some languages may exhibit more occurrences of high frequency phonemes than others which lead to lower SRTn scores. The highest average slope was found in the German and Polish DTT (17.1%/dB) whereas the lowest reported slope score was found in the Dutch version (10.2%/dB). When compared to the average normative value of the HINT test (10.3%/dB) (Soli & Wong, 2008), all MSTs except for the Dutch version showed higher slope of intelligibility.

As the Dutch version of the DTT had the lowest average slope score compared to the rest of the Hearcom's MST, Houben & Dreschler (2015) attempted to improve average slope by selecting sentences with steepest psychometric function to improve test sensitivity. Due to the design of the matrix sentence test, reshuffling of sentences is possible since all of the materials were equally evaluated and optimized. Producing new sets of test lists only requires re-evaluation of lists to determine changes in intelligibility function should the predicted average slope score has been adjusted. In this experiment, it was possible to improve the slope of intelligibility by preselecting sentences with high intelligibility scores however the authors cautioned that by altering the lists, the word occurrence and phonemic balance will be disrupted. This could cause the between lists sentences to be more familiar to listeners as some words and speech sounds may occur more than others.

To generate speech-shaped noise, most MSTs use their own speech materials which were superimposed by duplicating it many times. This allows the speech sounds to be masked effectively as the speech-shaped noise will have the same spectral content as the speech stimuli, leading to this being the recommended reference background noise for all MSTs (Kollmeier et al., 2015). The MST could also be used in different types of masker which makes the test more relevant to particular conditions of interest. Most commonly used in other tests are fluctuating noises, such as babble noise and/or temporally modified speech noise which is appropriate for making adjustments in hearing aids. They used fluctuating noises is known to reduce the SRTn in MSTs depending on the size of fluctuation or gaps between noises. Wagener & Brand (2005) investigated the effect of masker on the German MST (OLSA) in normal and hearing impaired listeners. Using two steady state speech shaped noises (summed speech material noise or test specific noise [TSN] & random Gaussian noise with male-weighted speech spectrum [icra1]) and two fluctuating noises (3-band speech fluctuating noise with male-weighted speech spectrum [icra5] & 6-talker babble [icra7]). The male-weighted noise was used as the OLSA uses a male speaker as its stimuli. Results of this experiment showed improvements of up to 14 dB for normal hearing listeners when icra5 was used compared to TSN, and a marginal 3 dB improvement when it was compared to icra7. Improvement of SRTn was also seen in hearing impaired listeners but was significantly smaller. The large improvement in SRTn when using the icra5 noise could be contributed to large pauses between noises (up to 2 seconds) that could have allowed listeners to capture part or whole sentences. Using fluctuating noise however showed larger variations between subjects especially in the hearing impaired group with some listeners either showed improvement or reduced performance. Using steady state noises showed more consistency in the test-retest examination of the OLSA. Hence it was recommended that for an adaptive measurement of SRT in noise to differentiate between listeners, a masking noise with spectral properties that are equivalent to the LTASS should be used. The MSTs in German, Russian and Spanish were also examined using different maskers in 10 native speakers of the language each (Hochmuth, Kollmeier, Brand, & Jürgens, 2015). In this study 9 maskers were studied, amongst a few the TSN, icra1, gated noise with short silent gaps were used (icra5-250) and multi-talker babble noise (MTN). The results were as expected where the listeners performed significantly better in icra5-250 compared to TSN and the benefit from release of masking was consistent across languages. Using MTN showed detrimental effect to listeners SRTn across languages from as large as 3.4 dB (Spanish MST) to as small as 0.2 dB (Polish MST) when compared to TSN. The MTN consists of 20 different talkers and used English as a medium which did not match any of the test languages hence not providing informational masking. The reduction in SRT performance in babble can be attributed to different voice onset times and larger fluctuations between noise which could distract listeners and cause higher attentional load. Effects of using MTN vary across languages when compared to both TSN and icra1. The Russian MST showed SRTn reduction of less than 1 dB when compared to TSN whereas the SRTn scores dropped by 4.5 dB when compared to icra1. This difference in response could be due to the differences in spectral content of TSN in Russian compared to the MTN. Note that the pattern of results were obtained in MSTs using bilinguals as speakers (Hochmuth, Jürgens, et al., 2015)

Based on the Hochmuth et al. (2015) and Wagener & Brand (2005) studies, it is clear that although masker type affected the results of the test, it could not explain the differences between various MST tests as they exhibited consistent results. To better understand the effects of the language, speaker and speech stimuli, Hochmuth, Jürgens, Brand, & Kollmeier (2015) recruited German/Russian and German/Spanish bilingual talkers (BT) to explore possible differences between these elements. First, words spoken by BTs for all four of these MST languages were re-recorded and optimized. It was then tested on native speakers of the languages to observe differences. In terms of syntactic structure of the tests, no differences were observed between languages. The BTs were identified their level of accentedness as it was expected that BTs with clear foreign accent could cause scores to be lower than their counterparts with no foreign accent. In this study, this was not observed as scores obtained were consistent between BTs who had clear foreign accent and BTs with no foreign accent with some BTs with no foreign accent resulting in lower SRTn. Spectral analysis between languages showed that they were similar and it could not have contributed to the differences in SRTn. Speaking rate were also observed and no differences were found between tests that had faster or slower speaking rates. Hence, for the MSTs other factors other than accentedness, spectral content and speaking rate could have affected inter-talker SRT differences. They concluded that the inter-talker differences found in the MST are mostly influenced by talker-specific differences instead of language or test design specific influences.

For future comparison and standardization of MSTs in various languages using multiple masking noises and/or speakers, a predictive model using voice recognition has been successfully tested using the OLSA to examine if an automatic process could be used (Schädler et al., 2015). Using this predictive model, it is possible that certain aspects of MST development can be relaxed by using simulation. The most exciting prospect of this model is the ability to predict performances of individual listeners based on their hearing levels which allows easier identification and (to some degree) explanation of a possible abnormal result.

In terms of non–native listeners ability to respond to the MST, Warzybok, Brand, Wagener, & Kollmeier (2015) compared performances of non-native German listeners using the German DTT, OLSA and the Gottingen sentence test. The results were consistent with other studies where non-native listeners performed significantly lower than native speakers of the language for the OLSA and Gottingen sentence test by 3 and 6 dB respectively. However, they showed native speaker like responses when tested using the German DTT. This is mostly due to the differences in the complexity of the task and the level of vocabulary used between the OLSA and the German DTT.

Test validation is a method to ensure that the test is able to do what it was designed for. So in the case of an audiological tool for speech perception in noise such as the MST, it can be validated by evaluating the responses of listeners with varying hearing levels and also by comparing results to other comparable established and standardized test. Jansen et al. (2012) compared the French MST to an everyday sentence tests in French called FIST and the French DTT in a group of 118 listeners with normal and impaired hearing. The French MST and FIST showed good positive linear correlation for both normal and hearing impaired listeners. Regression analysis between the tests showed the steep slope exhibited in the FIST showed lower discriminative power compared to the French MST. The French MST showed a significantly strong positive correlation to the French DTT which suggests that the two tests work well together as a screening tool and a clinical diagnostic assessment. The Turkish MST was compared to the Turkish HINT for validation purposes and was found to have higher slope of intelligibility at the respective levels of SRTn (Zokoll et al., 2015).

The structure of the matrix sentence test is flexible so it could also be designed to test children by reducing the number of possible combination of words making the sentences shorter with limited vocabulary so that it is easy for children to understand and respond to. An example of this is the Polish Paediatric MST designed for children age 3-6 years old (Ozimek et al., 2012). The sentence structure was reduced to [*subject* + *verb* + *object*] with 4 words per syntactic category. To increase the possible number of choices, 4 sub-matrixes were created to make the total usable sentence to 256 sentences (4^3 * 4 sub-matrixes). Instead of using a table of words for response selection, this test used a picture-point method with 256 different pictures which corresponds with each sentence created. For each listening task a six picture array is presented that has associated pictures as alternatives to compliment the correct one. The investigation of normative values showed an age effect between listeners. This led to a normative range between -1.0 to -2.9 dB for the target age group using picture pointing task. Children with hearing impairment showed significantly higher SRTn compared to their normal hearing counterpart. The German version of the paediatric MST called the Oldenburger Kinder-Satztest (OlKiSa) also was designed for preschool children but uses a pseudo-sentence structure of [numeral + adjective + object]. Test format selected is open set response where children will need to repeat what was heard. Similar observation of age dependent result was found for this test children in 1^{st} year of German primary school scored 1-2 dB higher compared to 2^{nd}, 3^{rd} and 4^{th} years children.

2.6.5 New Zealand English Auditory-Visual MST (McClelland, 2015; O’Beirne et al., 2015; Trounson & O’Beirne, 2012)

A team of postgraduate students led by Assoc. Prof Greg O’Beirne has successfully developed and optimized a New Zealand English MST that incorporates visual stimuli which allows the test to be conducted in auditory only, visual only and auditory-visual modes. As described in the previous section of the literature review, speech perception is ultimately multimodal and listeners use all the information (both visual and auditory) to process speech information. For example, speech can be as much as 50% more intelligible if they can see the face of the talker especially in low signal-to-noise levels (Sumby & Pollack, 1954). Visual information plays a huge role in providing speech cues to the extent that we can better resolve visual speech information if visemes are adequately redundant within a sentence (Tye-Murray, Sommers, & Spehar, 2007b). Taking into consideration the advantage in the flexible design of the MST and the importance of visual stimuli in speech perception, this led the team at the University of Canterbury to pursue an auditory-visual version of the MST (UC AV-MST).

In addition to the task of selecting suitable words and sentence structure, the design of the AV-MST requires additional consideration in recording and editing the materials. Trounson (2012) recorded both the video and audio recordings concurrently in an audiometric cabin using a high definition camera. He designed an custom AV recording segmentation rule using editing techniques adapted from Wagener et al. (2003) where between words co-articulation are preserved and additional attention was given to ensure the video segments are smooth between frames. This has led to some AV recordings to be edited within words as it seems to be the best possible ‘in’ and ‘out’ points to ensure smooth video transitions.

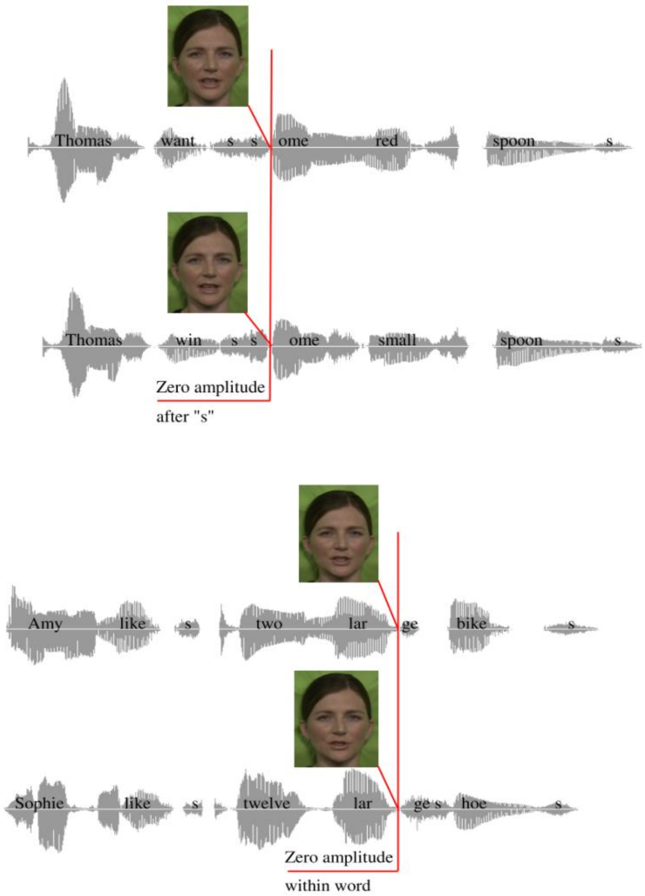


Figure 8: Example of auditory-visual segmentation used in Trounson (2012).

The testing interface software was designed by Assoc. Prof. Greg O’Beirne using LabVIEW. An analysis of the video transitions showed a substantial amount of video transition that was shifted in multiple axes between frames. Further analysis was done to exclude non-natural video frame transitions. Trounson suggested that various techniques to improve the video transitions but ultimately suggested a mechanical support during recording could possibly be the best option for future recordings. An important issue that was raised in his thesis was the possible scoring method to be used, as recordings are not whole words but fragments or parts of words that were segmented. The scoring method was adapted in a follow up study to normalise and refine the UC AV-MST (McClelland, 2015).

McClelland (2015) worked on two aspects of the UC AV-MST: (1) to identify and discard AV stimuli that contained perceivably noticeable judders; and (2) normalise audio recordings of the UC AV-MST. For her study (1), 18 native speakers of New Zealand English adult speakers with normal hearing were recruited. They were asked to rate if selected video sequences that contained varying levels of judders on a 10 point scale, from 0 (no noticeable judder) to 10 (highly noticeable judder). She found significant differences between ratings of transitions that contained no judder with transitions that were synthesized with judders. The decision was made include several synthesized transitions using a regression analysis so that the test could have adequate usable sentences within a test lists. For study (2), 17 adults with the same criteria as study (1) were recruited. The test was optimized using two noises: a test specific noise and a six-talker babble noise. An additional study was done to evaluate the scoring methods modified from those proposed in Trounson (2012), which is shown in Figure 9 below.

Actual	*Amy*		*bought*		*some*		*red*		*coats*		**Fragment scoring**		
Selected	*Amy*		*bought*		*some*		*red*		*coats*		**Pt1**	**Pt2**	**Total**
amy_bought	A	my	_	_							2/2	0/0	1
bought_some			bou	ght	s	_					2/2	1/1	1
some_red					_	ome	re	_			1/1	1/1	1
red_coats							_	d	co	ats	1/1	2/2	1
Word scoring	1		1		1		1		1				

Actual	*William*		*kept*		*ten*		*good*		*toys*		**Fragment scoring**		
Selected	*William*		*kept*		*ten*		*green*		*toys*		**Pt1**	**Pt2**	**Total**
william_kept	Will	iam	_	_							2/2	0/0	1
kept_ten			ke	pt	_	_					2/2	0/0	1
ten_good					te	n	_	_			2/2	0/0	1
good_toys							go	od	to	ys	0/2	2/2	0.5
Word scoring	1		1		1		0		1				

Actual	*Peter*		*has*		*six*		*good*		*mugs*		**Fragment scoring**		
Selected	*Peter*		*has*		*some*		*good*		*mugs*		**Pt1**	**Pt2**	**Total**
peter_has	Pe	ter	_	_							2/2	0/0	1
has_six			ha	s	s	_					2/2	0/1	0.667
six_good					_	ix	_	_			0/1	0/0	0
good_mugs							go	od	mu	gs	2/2	2/2	1
Word scoring	1		1		0		1		1				

Actual	*Oscar*		*got*		*four*		*red*		*books*		**Fragment scoring**		
Selected	*Oscar*		*got*		*four*		*red*		*books*		**Pt1**	**Pt2**	**Total**
oscar_got	Os	car	_	_							2/2	0/0	1
got_four			go	t	f	_					2/2	1/1	1
four_red					_	our	re	_			1/1	1/1	1
red_books							_	d	bo	oks	1/1	0/2	0.333
Word scoring	1		1		1		1		0				

Actual	*Hannah*		*sees*		*nine*		*large*		*toys*		**Fragment scoring**		
Selected	*Amy*		*sees*		*eight*		*cheap*		*spoons*		**Pt1**	**Pt2**	**Total**
hannah_sees	Han	nah	s	_							0/2	1/1	0.333
sees_nine			_	ees	_	_					1/1	0/0	1
nine_large					ni	ne	lar	_			0/2	0/1	0
large_toys							_	ge	to	ys	0/1	0/2	0
Word scoring	0		1		0		0		0				

Figure 9: Scoring procedure for words and fragments in the UC AV MST (Adapted from McClelland, 2015).

Results from the comparison between the normalisation using word or fragment scoring showed that slope scores were steeper using word scoring method hence further analysis of the AV-MST will use only the word scoring approach. The normalisation of audio stimuli showed a predicted SRTn and slope scores of -14.0 dB and 13.9%/dB for the test specific noise and -14.9 dB and 9.4%/dB for the six-talker babble noise. Participants scored slightly better in babble compared to the speech-shaped noise most likely due to release from masking. Based on the predicted values, 30 lists of 20 sentences were generated manually for each type masker for the AV-MST. Using these same predicted values, new lists were generated using an automated interative method by O'Beirne and Stone, resulting in two sets of 16 sub-lists of 10 sentences (one set for each noise type) designed to be randomly paired during testing (Stone, 2016).

2.6.6 Summary on MST

It is clear from reports above that the MSTs in various languages and noise configurations are able to produce accurate and reliable results, shown by their high slopes of intelligibility and small standard deviations in SRTn measurements. This suggests that MSTs have good discriminative ability and accuracy. Due to its design of semantically unpredictable sentences in fixed structures, the MST can be repeated in the same listener with no influences from learning or memorizing the materials as long as measures are taken, such as i) ensuring two training lists before commencement of recording SRT values; and ii) by using dual-track adaptive approaches to measure the SRTn and slope concurrently. The MST is appropriate for investigating speech perception abilities in noise or quiet for all hearing levels and is well suited for audiological evaluations that require repeated measure of speech performance, such as hearing aid or cochlear implant evaluations. A Dutch MST specifically optimized for cochlear implant users was developed for this purpose (Theelen-van den Hoek, Houben, & Dreschler, 2014).

Based on the studies reviewed above, it is known that masker type affects the results of the MST. Test specific noise or stationary broadband noises that have similar spectral characteristics of the long term average speech spectrum provide for optimal measurement of SRT in noise (Hochmuth, Kollmeier, et al., 2015; Kollmeier et al., 2015). Using temporally-modified gated noise would improve SRTn depending on the duration of silent gaps in the noise. For example, a 2 second gap in Wagener & Brand (2005) using icra5 noise showed an improvement of about 20 dB whereas the a 250 millisecond gap using a modified version of the same noise could improve SRTn by about 10 dB (Hochmuth, Kollmeier, et al., 2015; Zokoll et al., 2015). There were differences in SRTn and slope when various MSTs were compared between languages, and evidence suggests that this was mainly influenced by speaker-specific differences rather than differences in languages or the method of MST development (Hochmuth, Jürgens, et al., 2015; Kollmeier et al., 2015). Future development and comparison of new versions of MST should adapt methods suggested by the International Collegium of Rehabilitative Audiology (Akeroyd et al., 2015) using test specific noise as maskers.

Various versions of the MST (including paediatric versions and an auditory-visual MST in New Zealand English) have been developed as researchers found the matrix sentence test format to be applicable to different groups of listeners. Work is ongoing to develop an automatic prediction model to help diagnostic decisions in the future.

2.7 Introduction to the Malay language

Malay belongs to the Nuclear Malayo-Polynesian branch of Austronesian languages. Malay is spoken by people (mostly of Malay ethnicity) who live in the Malay Peninsula, Southern Thailand, the Philippines, Singapore, Central Eastern Sumatra, the Riau Islands, and part of the coast of Borneo (Asmah, 1993). It is an official language in Malaysia, Brunei, and Indonesia, and is one of the national languages spoken in Singapore. There are about 200 million speakers of this language worldwide.

In Malaysia, Malay is known as Bahasa Melayu with 16.5 million speakers. In Indonesia, it is known as Bahasa Indonesia (170 million speakers), in Brunei and Singapore, the variants are known as Bahasa Melayu Brunei (250,000 speakers) and Bahasa Melayu Singapura (3.25 million speakers) respectively (Asmah, 1993). There are also many versions of colloquial Malay spoken by specific sub-ethnic Malay groups and also the use of "Manglish" which is a blend of English and Malay which is used by Malaysian bilinguals with English education background (Ozog, 1987).

The main influence on the Malay language is Sanskrit followed by Arabic and English. The endurance of the Malay language is partly due its ability to adapt and adopt words from major languages that were widely spoken in trade, missionary activity and foreign occupation (Benjamin, 2009). The written Malay is now most commonly used in *Romanized* form while the *Arabic* form is used almost exclusively for the teachings of Islam.

2.7.1 The Malay sounds, morphology and syntax

The Malay phonology consists of six short vowel sounds (Figure 10), three diphthongs (Figure 11) and twenty-one consonant sounds (Table 4).

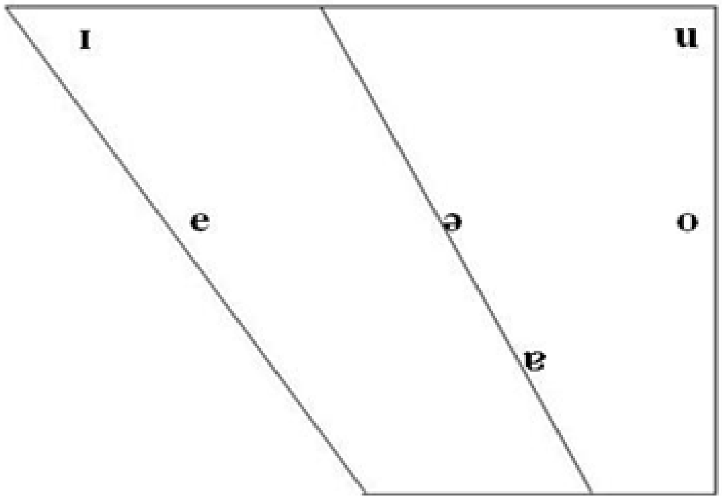


Figure 10: The vowel phonemes of standard Malay (Wan Aslynn, 2012)

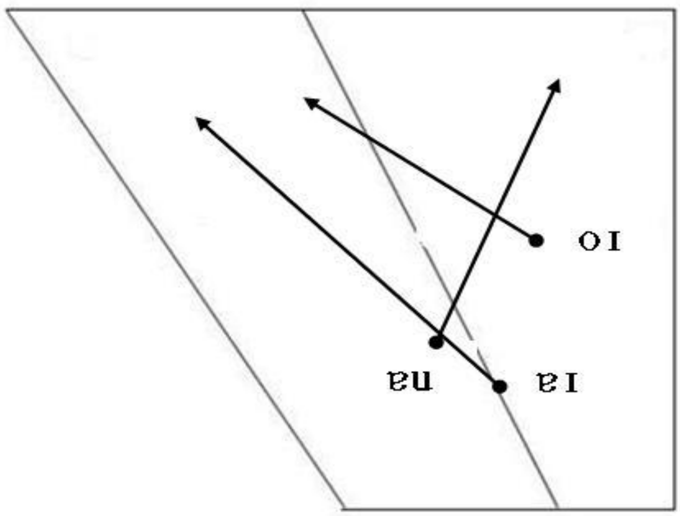


Figure 11: The diphthong phonemes of standard Malay (Wan Aslynn, 2012)

Table 5: The consonant phonemes of standard Malay (Yunus, 1980)

	Bilabial	**Labio-dental**	**Dental**	**Alveolar**	**Post-alveolar**	**Palatal**	**Velar**	**Glottal**
Nasal	m			n		ɲ	ŋ	
Plosive	p b			t d			k g	
Affricate					ʧ ʤ			
Fricative		f v		s z	ʃ			h
Approximant				ɹ		J		
Lateral				l				
Other consonant: w								

Malay is classified as an agglutinative language in which words are formed by joining morphemes together (Wan Aslynn, 2012). Generally there are two types of words in Malay; morphologically simple words and morphologically complex words. There are three types of word formation found in morphologically complex lexical items, namely affixation, compounding and reduplication.

There are three types of affixes permissible in Malay, namely prefixes, suffixes and circumfixes or infixes. Some of the prefixes and suffixes can change the word class of a stem. For example, the prefix '*pel*' attached to a verb stem will produce a noun, so attaching *pel* to the stem word *ajar* 'to teach' (a verb) creates the noun 'pelajar' (student). Similarly the suffix '*-an*' if attached to the end of the stem alters the word class, for example the noun '*ajaran'* (teachings). Other affixes such as prefix '*bel'* and suffix '*-kan*' will not change the word class of the stem, for instance, the verbs '*belajar'* (to learn) and '*ajarkan'* (to teach).

The derivation of new words by infixes will always change the word class. Malay allows attachment of more than one affixes in infixes, for example the word ‘*pembelajaran’* ‘the process of learning’ (noun) is derived from prefix ‘*pem*-’and infix ‘*bel*-’and suffix ‘-*an*’, they are attached to the verb stem ‘*ajar’* (to teach).

New words can also be formed by the combination of two root words and is restricted to the combination of the same word class (noun-noun or verb-verb). The process of compounding does not change the word class of the newly combined words, for example the word *kereta* ‘car’ (noun) when combined with the word *api* ‘fire’ (noun) will become *keretapi* ‘train’ (noun). Affixation also is permissible on compound words, and this sometimes can alter the word class, for example the prefix ‘*peng-’* and suffix ‘-*an*’ can be added to the word *ambil alih* which means ‘to take over’ (verb) – seperately *ambil* means ‘to take’ (verb) and *alih* ‘to move’ (verb) – to become *pengambilalihan* ‘the process of taking over’ (noun).

Another way to form a new morpheme in Malay is reduplication. According to Chang (n.d) reduplication usually is used to indicate plurality. For example the word *kotak* ‘one box’ if duplicated becomes *kotak-kotak* ‘more than one box’. If the reduplication process happens on a verb, it alters the word class to a noun, for example the word *main* ‘to play’ (verb) if duplicated becomes *main-main* ‘banter’ (noun).

Malay words are classified into eight classes or parts of speech – namely nouns, pronouns, verbs, adjectives, adverbs, conjunctions, prepositions, and interjection and the most typical word order is Subject-Verb-Object (SVO) (Asmah, 1974). Typically, the noun phrase (NP) is immediately followed by the verb phrase (VP). In Malay the sentence may be simple or compound. The simple sentence contains only one clause and the compound sentence contains more than one clause. The compound sentence uses conjunction to support the sentence structure. Coordinating conjunctions like *“dan”* (and) or *“atau”* (or) and sub-coordinating conjunctions like *“kerana”* (because), *“jikalau”* (if), *“sungguhpun”* (although) *“bahawa”* (that) are typically used in written Malay but are not frequent in dialectal Malay.

Examples of simple sentences:

a. Dia memandu kereta

‘He/She + drive + car’ or “He/She drives a car”

b. Dia sedang mencari seorang doktor yang kaya.

‘He/She + is + looking + a + doctor + the + rich’ or “He/She s looking for the rich doctor”

The following is an example of a compound sentence taken from a published editorial (12th October 2012) from Utusan Malaysia, a mainstream Malay daily:

“Penerokaan angkasa lepas demi kebaikan manusia sejagat sama ada dari segi ekonomi atau saintifik memerlukan pelaburan yang besar dan jangka masa panjang untuk membuahkan hasilnya yang bakal diperoleh pada masa hadapan”.

“Space exploration requires large investments and extensive periods to produce the results that will be obtained in the future for the benefit of mankind either economically or scientifically”.

2.7.2 Malay matrix sentence structure

Generally, before we are able to develop a list of matrix sentences, it is important to know the principles that underlie the language. There are rules that govern the internal part of language – namely phonetics, phonology, morphology, syntax, semantics and pragmatics. Each principle carries its own function in creating the identity of the language. The table below summarizes the sentence structure and source of materials of other currently available matrix sentence tests.

Table 6: Comparison of language structure in selected international versions of matrix sentence tests.

Matrix Sentence Tests	Source of Speech Materials	Sentence arrangement	Example of sentence
French Sentence Test (Jansen et al., 2010)	Based on corpus of spoken French	Noun + Verb + Number + Object +Adjective	French : Emile voudrait deux livres rouges English: Emile wants two red books

Spanish Matrix Sentence Test (Hochmuth et al., 2012)	The words were selected according to their frequency of usage in Spanish language	Noun + Verb + Number + Object + Adjective	Spanish : Claudia tiene dos libros grandes English : Claudia has two big books
The British English Matrix Test (Hall, 2006)		Noun + Verb + Number + Adjective + Object	Peter got three large desks
Danish Sentence test (Dantale II) (Wagener et al., 2003)	Written Danish sentence such as Dictionary of Word Frequency in Danish	Noun + Verb + Number + Adjective + Object	Danish : Anders ejer ti gamle jakker English : Anders owns ten old jackets
Polish sentence test (Ozimek, Kutzner, Sek, & Wicher, 2009)	Written Polish sentences from daily speech, literature, TV & theatre	Name + Verb + Number + Adjective + Object	Polish: Zofia wygra wiele czarnych okien English : Sophie will win many black windows
German Matrix Sentence Test (Wagener, Kuhnel, & Kollmeier, 1999)	Speech compilation of male unschooled speaker	Noun + Verb + Number + Adjective + Object	German : Kersti seiht neun kleine tassen English : Kerstin sees nine small cups
New Zealand Matrix Sentence test (Trounson & O'Beirne, 2012)	British matrix was used as basis.	Noun + Verb + Number + Adjective + Object	New Zealand English : Amy bought two big bikes

In order to develop a standard syntactically correct five words Malay sentence with equal language complexity, the following structure will be used;

Noun + Verb + Number + Object + Adjective

This order is similar to that found in the French and Spanish versions of the Matrix sentence test, as adjectives generally follow objects in those languages.

In written Malay, an object descriptor will usually be placed between a numerator/number and the object - for example "tiga *keping* cawan" or three *(pieces)* cups. The object descriptor (in this case, "*keping"*) is used to provide clues about the nature of the object. For example, the word "biji" is a descriptor that is normally used to objects that have a round figure such as a ball ("tiga *biji* bola") or an apple ("tiga *biji* epal*")* and the descriptor word "helai" is used to describe an object that is thin such as paper ("tiga *helai* kertas") or cloth ("tiga *helai* kain"). In its verbal form, the object descriptor is often used in formal communication such as in speeches or teaching-learning sessions. For the purpose of this study, the object descriptor will be omitted to reduce acoustical input and maintain homogeneity. This omission will not affect the overall meaning of the sentence and will be helpful to this test as it will remove some contextual clues of the nature of the object.

Referring to the suggested Malay sentence structure, the conjunction yang 'the' is also typically used between an object and adjective to accurately explain the context of a sentence. For example, "Siti ada tiga kasut *yang* cantik" (Siti + has + three + shoes + the + beautiful or 'Siti has the three beautiful shoes) which means Siti has many shoes and three of them are beautiful. By omitting the word "yang" in the previous example, the context is affected where it now means, Siti has only three shoes and they are beautiful. For the purpose of this study the conjunction "yang" will not be used as it only affects the context of the sentence and not the syntax. This is also to reduce acoustical redundancy within the matrix sentence test, as every sentence would contain that same conjunction.

2.8 Audiology & hearing loss in Malaysia

Audiology as a formal education was first introduced in Malaysia by the Universiti Kebangsaan Malaysia *(National University of Malaysia)*. The university initiated a four year bachelor's degree in audiology in 1995. Subsequently two other public universities – the International Islamic University of Malaysia and the Universiti Sains Malaysia (*Science University of Malaysia*) offered similar programmes at undergraduate level in 2004 and 2005, respectively. By 2010, there were some 90 audiologists in service in government hospitals with at least the same figure working in the private health sector, such as private hospitals and hearing aid dispensing centres (Ministry of Health Malaysia, 2012).

As the numbers of hearing professionals steadily increased since the early 2000s, there was an amplified interest in hearing awareness and its psychosocial implications. As a result, the first national survey on hearing and ear disorders in Malaysia was conducted in 2005 (Abu Bakar, 2007). The survey revealed that the prevalence of hearing impairment in Malaysia was 17.1%, 1% of which was identified to be profound hearing impairment. The nationwide study with a sample size of 7041 was conducted based on the World Health Organization (WHO) Ear and Hearing Disorders protocol with small adjustments made to meet local settings. Over 75% of the subjects found with hearing loss were adults aged over fifty, with more male subjects found with hearing impairment than females. At 5%, the prevalence of a significant or disabling hearing loss (moderate hearing loss or more) in Malaysia is similar to other South East Asian countries, which ranged from 4.2% to 9.1% using equivalent hearing screening protocols (WHO, 2007). As a result of this survey, the Ministry of Health Malaysia took several measures, such as initiating hearing awareness campaigns, implementation of high risk neonatal hearing screening in 30 hospitals, and also the introduction of a subsidized cochlear implant programme. These measures were designed to help create greater awareness as well as making resources more accessible to the public. The data from this study have also helped in the recognition of the need for a more comprehensive audiological service in all public hospitals.

2.9 Speech testing in Malaysia

Audiological services in Malaysia are available only major hospitals in each state with some private institutions offering diagnostic and hearing aid dispensing services. The diagnostic investigations are usually limited to pure tone and impedance audiometry. The audiogram is often the only source of information as far as subjective hearing sensitivity is concerned and consistencies of results are only observed between the audiogram, tympanograms and reflexes. Some centres may conduct auditory brainstem response (ABR) testing as a cross check if necessary especially for difficult to test children. Because of the limited resources and many languages and dialects spoken in the country, speech audiometry is usually not part of the routine investigation.

The first published attempt to standardize a Malay word list was done by Hong in 1984. In this study, a set of ten lists each containing ten disyllabic words were created and analysed by comparing results between native Malays and non-Malay speakers who are familiar with the language.

Table 7: List of disyllabic Malay words in Hong, (1984).

1	2	3	4	5
Lama/old	Bulan/moon	Lari/run	Laki/man	Kati/measure
Hari/day	Kiri/left	Surat/letter	Isi/fill	Sudah/done
Suka/like	Kata/say	Besi/iron	Juta/million	Tipu/lie
Gaji/pay	Apa/what	Muka/face	Cuka/vinegar	Tahu/know
Cuba/try	Hati/heart	Buta/blind	Jadi/become	Sama/equal
Mati/die	Sapu/wipe	Tidak/no	Budak/boy	Silat/silat
Buruk/ugly	Gula/sugar	Anak/son	Telah/after	Budi/deed
Alat/device	Cuci/clean	Roti/bread	Besar/big	Bila/when
Sakit/pain	Nasi/rice/	Sana/there	Diri/self	Atas/top
Batu/stone	Lupa/forget	Mahu/want	Akan/will	Jari/finger

6	7	8	9	10
Kami/us	Susah/trouble	Mula/start	Laku/valid	Mata/eyes
Cuma/just	Lagu/song	Tadi/just now	Masa/time	Rupa/looks
Ada/have	Dulu/then	Basah/wet	Maju/advance	Padi/paddy
Satu/one	Bisu/dmb	Kasut/shoe	Bumi/earth	Kalah/lose
Biji/piece	Tiba/arrive	Ahli/member	Tali/rope	Saya/I
Bukit/hill	Lima/five	Buku/book	Sijil/certificate	Sila/please
Mari/come	Buka/open	Gila/crazy	Bola/ball	Murid/student
Lagi/more	Jika/if	Jiwa/soul	Arak/alcoholic drink	Mesti/must
Biru/blue	Ajar/teach	Beli/buy	Kaki/feet	Juga/including
Tiga/three	Pasar/market	Cuti/holiday	Pagi/morning	Adek/brother

Several issues could be identified in this study; first, the author intended the test to be based on a phonemic scoring system and a full mark of 10% is given for each word. Further inspection of the word lists shows that some words included had more than four phonemes with unequal number of phonemes for each list. Secondly, the author did not mention the process of evaluating the words included in the list. This could lead to some bias either in the level of intelligibility or social acceptability. It is noted that the interlist comparison showed no significant difference but a trend of increased SRT was seen from list 1 to list 10. To validate the results, the author compared results between native Malay speakers and non-Malay subjects who are familiar with the language and found no significant differences. Some of the words used are questionable, words like “gila” which means crazy and “arak” which means alcoholic drinks has negative associations. In addition, the word “laki” which is supposed to mean man or men, is today no longer the accepted terminology and has been replaced by the word “lelaki”.

The most noted author for the development of speech tests in Malaysia is Professor Siti Zamratol Maisarah Mukari from the Universiti Kebangsaan Malaysia. Mukari & Said (1991) developed a phonemically balanced disyllabic Malay (c-v-c-v) word list to be used in speech audiometry which was based on the work by Hong (1984). Disyllabic word structure was used as it is the most commonly found word structure in Malay (Asmah, 1971) The researchers compiled 25 word lists and used 196 Malay words with some words repeated as much as five times between lists to maintain its balance phonemically. At the point of speech discrimination threshold (50%) the standard deviation of this test was 5.5 dB (22.3%) which is considerably higher that what was suggested by Suter (1985) to ensure clear difference between normal and impaired hearing. This test is not commonly used in clinics and the reasons for this are not clear. Further analysis of the words lists show that the content are only phonemically balanced between lists but was not referenced to any literature in terms of familiarity or frequency of phonemes.

Phoneme	List 1	List 2	List 3
m	1	1	0
p	1	1	1
b	1	1	1
f	0	0	0
v	0	0	0
n	1	1	1
t	1	2	1
d	1	1	1
s	2	1	2
z	0	0	0
r	1	1	1
l	2	2	2
ʧ	1	1	2
ʤ	1	1	1
ʃ	0	0	0
ɲ	0	1	1
J	1	1	1
ŋ	2	1	1
k	1	2	1
g	1	1	1

h	1	1	1
ʔ	0	0	0
w	1	0	1
ɪ	6	4	3
ə	4	4	6
ɛ	1	1	1
ɐ	5	5	5
u	4	5	4
o	2	1	1
ɐ ɪ	0	1	1
ɐ u	1	1	1
oɪ	0	0	0
sum	43	42	42

Table 8: Comparison between frequency of phoneme in word list 1, 2 and 3 (Mukari & Said, 1991)

Other published work on speech testing by Professor Mukari are; the development of single and double dichotic digit tests in Malay for the diagnosis of auditory processing disorders (Mukari et al., 2006) and the Malay Hearing in Noise Test (MalayHINT) (Quar et al., 2008). The MalayHINT was designed for the selection of hearing aids as well as for validation purposes.

Since a speech test using words in Malay is not yet standardized, Marina L. Alisaputri (a Malaysian PhD candidate at the University of Sheffield, United Kingdom) is attempting to produce a single word speech test as part of her doctoral thesis (Marina Alisaputri, personal communication, February 11, 2016). At the moment she has compiled lists of disyllabic words in CVCV structure. The test was developed in two versions, (1) test with lists of meaningful and nonsense words and (2) test lists with meaningful words only. Test are scored phonemically similar to the AB words list test (Boothroyd, 1968). She found no statistical difference between the two versions of the test and psychometric evaluation of the test was found to have strong correlations to severe and profound hearing losses. She is expected to complete her studies by July 2016.

CHAPTER 3

DEVELOPMENT OF THE MALAY DIGIT TRIPLET AND THE AUDITORY-VISUAL MATRIX SENTENCE TESTS

3.1 Background

This study involves the development of two speech intelligibility tests which are:

- the digit triplet test in Malay using headphone and telephone (MDTT); and
- the Malay auditory-visual matrix sentence test (MMST-AV).

The development of both tests were done concurrently as it was efficient for both tests to go through similar processes using mostly the same participants at each level of the development and evaluation of the tests. After normalisation, the tests were incorporated into the University of Canterbury Adaptive Speech Test (UCAST) software platform developed by Associate Professor Greg O'Beirne.

In this chapter, a study to select an appropriate speaker for both tests is described, as speaking rate, articulation and voice quality are important factors in a speech test (Bradlow et al., 1996). This is followed by descriptions of the development of both tests, beginning with the selection of digits for the DTT and the selection of appropriate words for the Malay auditory-visual MST. The development of both tests involved recording and editing processes, as well as the creation of two background noises for each test. Also included in this chapter are the descriptions of the research apparatus used throughout this study.

Shown below is the summary of processes involved in the development and evaluation of the tests listed above.

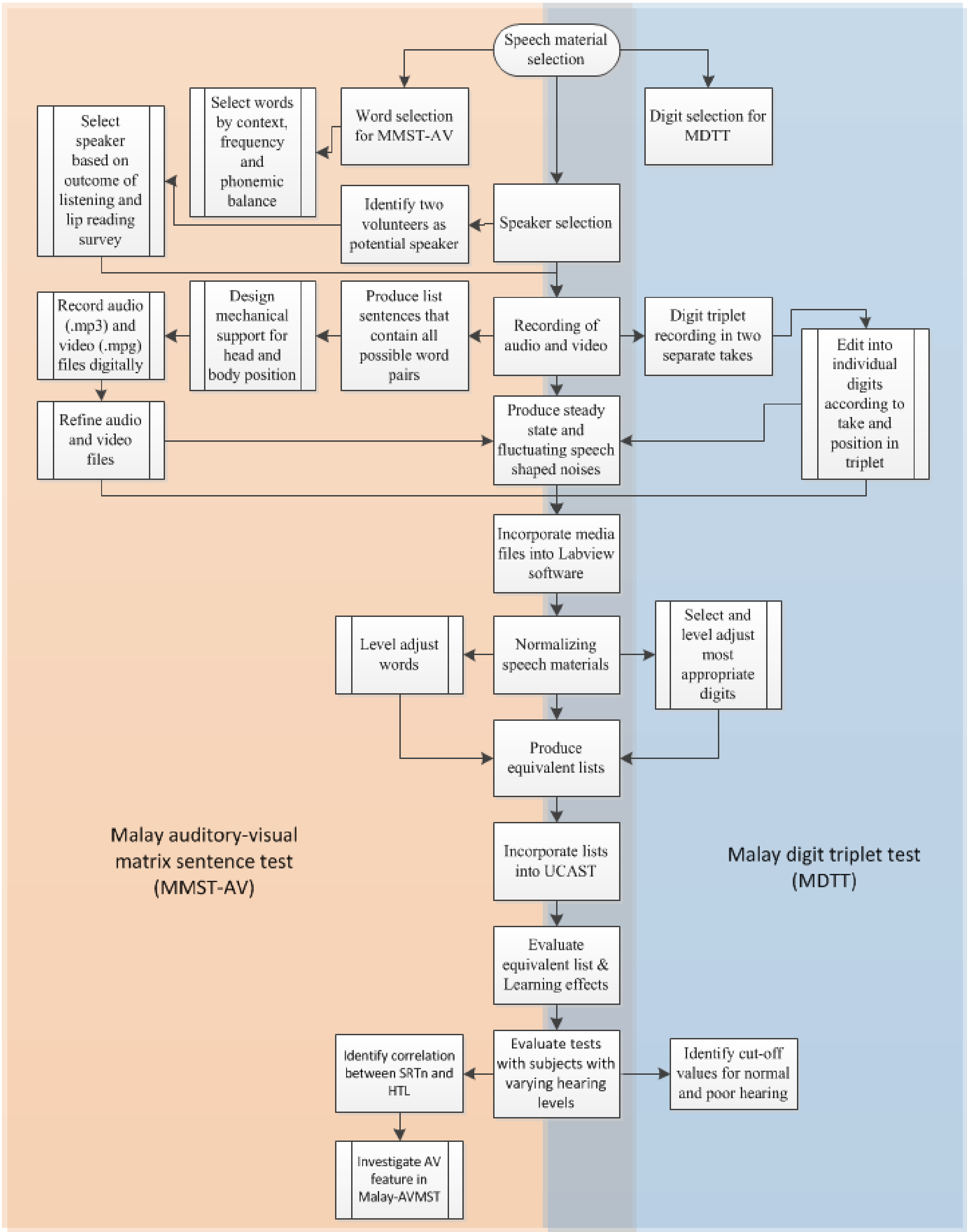


Figure 12: Processes involved in the development both tests

3.2 Speaker selection for both MDTT and MMST-AV

3.2.1 Introduction

As the Malay matrix test was designed to incorporate an auditory-visual testing mode, it was essential that the speaker for this test was able to provide clear and perceptible auditory and visual cues. Even if the same speaker changes vocal effort in the same sentences, it can have a significant effect on how well listeners can respond to the sentence (Mullennix et al., 1989). The objective of this part of the study was to identify an appropriate untrained speaker that would provide the voice for the digit recordings and sentences in the Malay matrix, as well as feature in the video portion of the Malay matrix test. As the MMST-AV involves an auditory-visual component, special attention was given to select a speaker that was not too difficult to lipread.

3.2.2 Speaker selection

Two female native Malay speakers volunteered to participate as the potential speaker in the selection study. Both speakers used standard Malay (Melayu Riau) in their daily communication and both were audiologists currently pursuing post graduate degrees at the University of Canterbury and had no prior experience or training as professional presenters for this purpose.

Fifteen native Malay speakers (mean: 33.3 years, S.D ± 5.6) in Christchurch participated in this study. Participants were either postgraduate students or school teachers with a minimum of 3 years working experience. All participants reported that they had normal hearing. Before the study was conducted, participants were briefed on the purpose of this study and how to rate the audio and visual components of the study.

3.2.2.1 Development of listening and lip reading study for speaker selection

Two tasks were developed to compare the speakers' overall voice quality and lip readability. The two volunteers were asked to read 20 sentences taken randomly from the Malay matrix sentence test. The speakers were asked to speak using neutral and natural intonation at their normal speech rate using standard Malay.

To evaluate the speaker's voice quality, four areas were identified and measured using a three point Likert scale. The four areas were intonation, voice projection, clarity and speech rate. To evaluate the speakers' lip readability, the audio of 5 sentences was removed and sentences were randomly shuffled to avoid similar sentence order and memorizing effects between the two speakers. Both speakers were recorded in high definition (1080i) at 50 fields/second using a Sony Handycam (model HXR-MC50e) in a single walled audiometric room. Speakers were also asked to wear the same colour headscarf to avoid any distractions.

Figure 13: Screenshots of the stimuli for the lip reading task

3.2.2.2 Rating the speakers

For the first part of this study, participants were asked to listen to five sentences without any visual cues. Sound intensity was set by each participant at a clear and comfortable level. For the speaker's intonation, voice projection and clarity, participants rated them "poor", "fair", or "good". Speech rate was determined by participants as either "good", "too fast", or "too slow".

For the lip readability study, participants were asked to write down the sentences that were spoken by the two speakers by lip reading. Adequate time was given between each sentence to allow participants to think or guess. The score was determined by comparing total correct words lip read by participants between the two speakers. The overall score between the two speakers was to be compared and the speaker with the highest score selected as the speaker for the Malay digit triplet test and the Malay matrix sentence test.

3.2.3 Results and analysis

Speaker 1 had an average speaking rate of 3.6 syllables per second whereas Speaker 2 had a faster average speaking rate of 4.4 syllables per second. Participants rated Speaker 1 as being either good or too slow and Speaker 2 rated as good or too fast with an average score for speaking rate of 2.5 and 2.9 respectively.

Table 9: Scoring differences between Speaker 1 and Speaker 2

	Voice quality score for speaker 1	Lip readability score for speaker 1	Total score for speaker 1	Voice quality score for speaker 2	Lip readability score for speaker 2	Total score for speaker 1
Mean	13.86	6.60	20.46	14.13	3.93	18.06
Median	14.00	7.00	22.00	15.00	4.00	18.00
σ	1.12	3.73	4.21	1.06	2.52	2.60

A paired sample T test showed no significant difference in voice quality scores ($t(14) = -0.718$, $p = 0.48$) between speakers. However, a significant difference was observed in both lip readability scores ($t(14) = 4.18$, $p < 0.05$) and total score ($t(14) = 2.8$, $p <$

0.0%) between speakers that suggested that participants found that the first speaker was easier to lip read than speaker 2.

3.2.4 Discussion

Visual information can contribute as much as 50% towards intelligibility and dependence on visual cues is more important at low signal-to-noise ratios because listeners may not benefit from auditory presentations in adverse listening conditions, hence supplementary visual observation of the speaker is recommended (Sumby & Pollack, 1954). The contribution of visual information comes from understanding and predicting lip movements, and this is evidently a top-down process through identification of visual lexicons (Tye-Murray et al., 2007b). In that study, sentences with more frequent visemes were more easily identifiable even when informational masking was applied. Based on the result of that study and the importance of ensuring visual cues are not lost during the MMST-AV, speaker 1 was selected to contribute the audio visual input for both digit triplet test and the matrix sentence test in Malay.

3.3 Digit selection for the Malay digit triplet test

3.3.1 Introduction

The aim of this part of the study was to identify appropriate digits in Malay to be used in the MDTT.

Shown below are digits that were selected by previous researchers to develop DTT versions in different languages. Most researchers choose digits that contain the same number of syllables to ensure that they are equally perceptible. This is however not the case for the Polish DTT as the number of mono and disyllabic digits in Polish are almost the same, hence both mono and disyllabic digits were used.

Table 10: Digits used in other versions of digit triplet test

Language	Digits used	References
Dutch	0, 1, 2, 3, 4, 5, 6 & 8 (monosyllabic)	Smits et al., 2004
New Zealand English	1, 2, 3, 4, 5, 6, 8, 9 (monosyllabic)	King, 2010
Te Reo	0, 1, 2, 3, 5, 6, 7, 8 & 9 (disyllabic)	Murray, 2012
Polish	0-9 (mono and disyllabic)	Ozimek, Kutzner, Sek, & Wicher, 2008
German	0, 1, 2, 3, 4, 5, 6, 8 & 9	Wagener et al., 2005

The inclusion criteria for digit selection for the MDTT are listed below;

i. Digits in Malay

ii. Disyllabic digits only

The Malay digits selected for recording are the digits 0, 1, 2, 3, 4, 5, 6, 7 and 8 as these digits are disyllabic. The three syllable digit 9 or “sembilan” was omitted to maintain homogeneity. Translations of the digits used are shown below;

Table 11: English and IPA translation of digits used in the Malay digit triplet recording.

English	Malay	IPA Transcription
Zero	Kosong	/kosoŋ/
One	Satu	/sɐ tu/
Two	Dua	/duə/
Three	Tiga	/tIgə/
Four	Empat	/əmpɐ t/
Five	Lima	/lImɐ /
Six	Enam	/ənɐ m/
Seven	Tujuh	/tuʤ oh/
Eight	Lapan	/lɐ pɐ n/

3.3.2 Discussion

The Polish version of DTT used both mono and disyllabic digits as it would be left with only 4 digits if the disyllabic ones were excluded (Ozimek, Kutzner, Sęk, et al., 2009). Selection of either mono or disyllabic digits would cause the number of possible triplets and triplet lists to be too low. The mean SRTn using the combination of mono and disyllabic digits in the Polish digit triplet test was found to be closely equivalent to the German and English versions, suggesting no influence of number of syllable was found in the Polish DTT. This is because precautions were made to ensure that the lists used were statistically and phonemically equal. Number of syllables is not an issue in the Malay version of the digit triplet as the digits chosen are all disyllabic. In later stages of development, the lists of triplets will contain equal frequency of digits.

3.4 Development of the Malay digit triplet test (MDTT)

3.4.1 Introduction

The aim of this study was to record, edit and generate digit triplets in Malay for the purpose of normalisation. The methods were based on previous work by King (2010) , Smits et al. (2004) and Zokoll et al. (2012).

3.4.1.1 Recording of speech materials

A female native Malay speaker (aged 33) was selected based on the result of the speaker selection study. The selected digits were recorded in mono in a single walled audiometric cabin at a sample rate of 44.1 kHz and saved in the ".wav" file using a TEAC TASCAM (model HD-P2 Portable High Definition Stereo Recorder). The speaker was asked to read three lists containing Malay digit triplets twice. The lists contained triplet combinations with all nine selected digits (0, 1, 2, 3, 4, 5, 6, 7, and 8) at all three positions. The repeat recording was to ensure the best recording for each digit was obtained. The best two read lists were determined by the author and were thereafter identified as Take 1 and Take 2 to be used in the normalisation process. Both Take 1 and 2 were then edited to separate audio files for each digit at each position. An additional 50 milliseconds of silence was then added before and after each audio file using the Audacity™ software to ensure noise bursts are eliminated between digit transitions. Additionally, several recordings of the word "nombor" (number) were made as the designated carrier phrase before each triplet is presented, and the "best" recording was chosen by the author. A playlist of 180 triplets each for headphone and telephone normalisation was created that selected two recordings (Takes 1 and 2) of all 9 selected digits at all three digit positions. All triplets with either two or three consecutive presentations of the same digit were excluded from this playlist. The "nombor" carrier phrase was added at the beginning of each triplet.

3.4.1.2 Development of masking noises of Malay digit triplet test

As described earlier in the literature review, the DTT has some limitation in discriminating between normal hearing listeners and listeners with highly sloping hearing loss. The design of a spectrally and temporally modified speech noise in the Telscreen II had shown some potential in increasing the sensitivity of the test (Golding et al., 2007). Therefore two types of noise were designed for the MDTT. The aim of producing these two noises was to investigate which masking noise could increase the specificity of the digit triplet test. The first noise has been typically used in other digit triplet tests, namely the test specific noise or also referred to by other authors as steady state speech-shaped noise. Also tested was a novel modulated speech-shaped noise which was designed to achieve optimum masking release in normal hearing listeners, referred to as spectrotemporal gap noise (STG noise).

3.4.1.3 Test specific noise

A speech shaped masking noise was generated by superimposing all recorded Malay digits 10 000 times within a 10 second looped sound file using an automated process. This produces a noise with identical spectral components to the speech stimuli, which effectively reduces the effects of using different transducers (Smits et al., 2004), as long as external noise remains within reasonable low levels. During testing, the presentation level for the test specific noise for the Malay DTT was set at a constant level (in dB A), while the level of the triplet stimuli was varied to achieve the desired SNR.

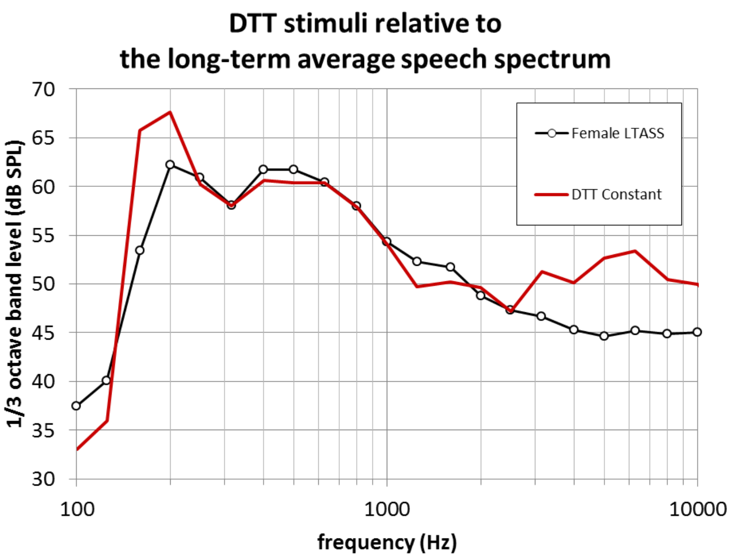


Figure 14: Spectra of the signal and speech shaped noise (overlying) used in the MDTT compared to the female long term average speech spectrum (data from Byrne et al., 1994).

3.4.1.4 Modified speech-shaped noise with spectral and temporal gap (STG) modulation

The second noise was produced by modifying the speech shaped noise by adding spectral and temporal gaps between signals. The STG noise is generated by creating two separate speech noise files with opposite temporal gaps by multiplying the speech noise sample with a 16 Hz trapezoid (10% rise-fall time) or with the opposite function (i.e. one that is 180 degrees out of phase). The two resulting noise files had complementary 16 Hz temporal gaps that were 100% modulated, such that addition of the two waveforms resulted in the original unmodified file.

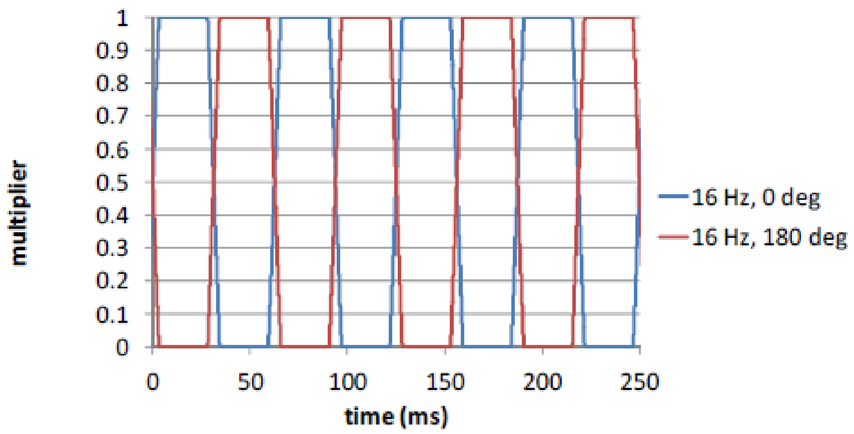


Figure 15: Multiplier waveforms that were used to create the temporal gaps (Adapted from Bowden, 2013)

Spectral gaps were introduced to the two noise files by multiple band-pass filtering. The gaps were two equivalent rectangular bandwidths (ERBs) wide (Glasberg & Moore, 1990), a bandwidth chosen as it was found to be unfavourable to hearing impaired listeners (Peters et al., 1998).

The ringing introduced by this filtering resulted in the 100% modulation produced by the temporal gap process reducing to a modulation depth of around 13 dB. Figure below shows a 250 ms excerpt of the two resulting modulated noise waveforms.

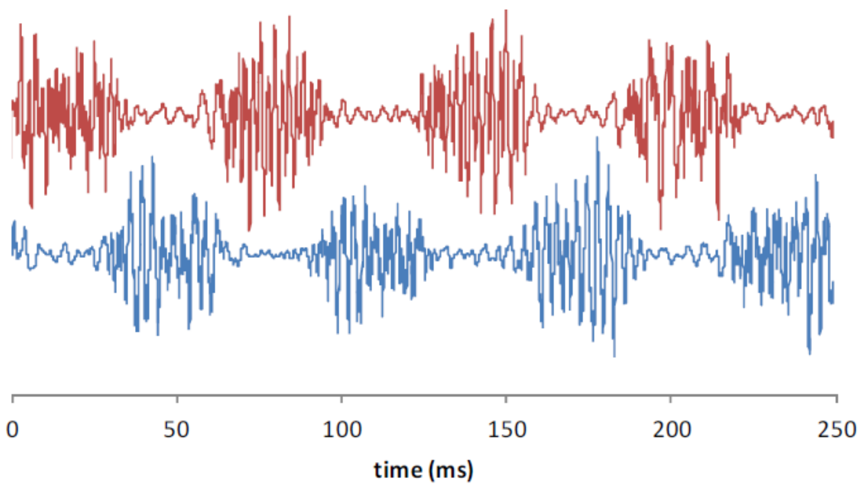


Figure 16: 250 milisecond excerpt of two resulting modulated noise waveforms, each of which has complementary spectral gaps (Adapted from Bowden, 2013).

The time domain structure is not affected even when the spectral content of the two waveforms are different, the interleaving waveforms cancels out the opposite gap (at the size of ERB) as both waveforms complement each other in opposite phases. The spectra of the two filtered waveforms are shown below. Therefore, the design of the STG noise allows for 2-ERB wide spectral gaps (at any given point in time) and 16 Hz temporal gaps (at any given frequency).

A.

B.

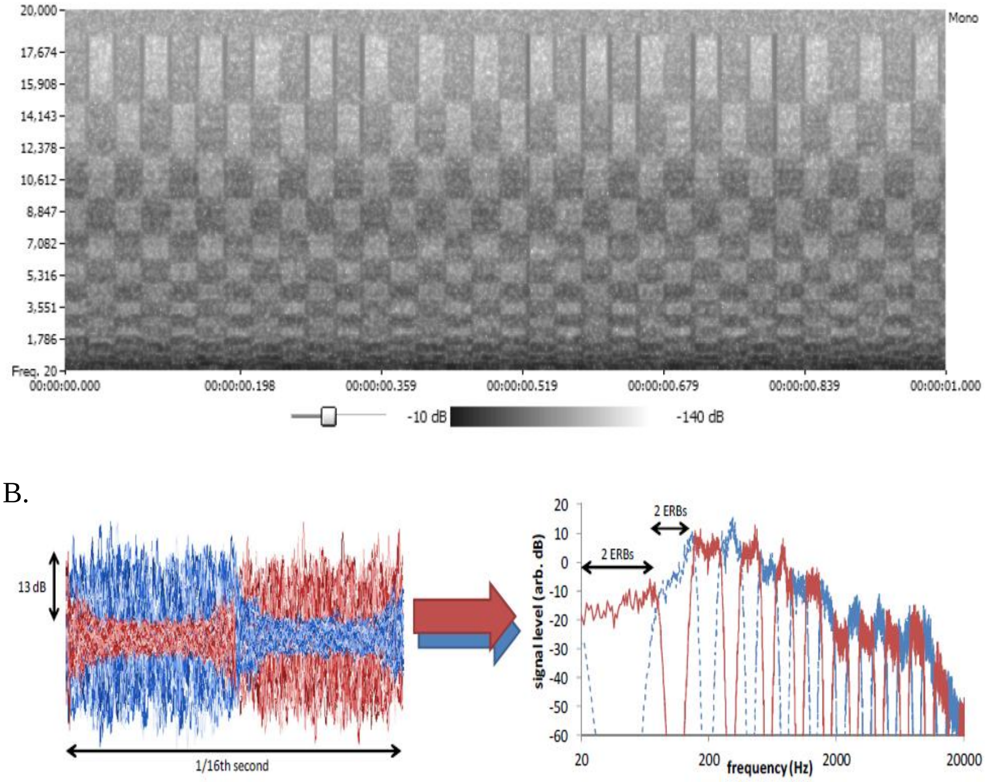


Figure 17: Spectral analysis of STG noise used in MDTT. (A.) A spectrogram of the STG noise showing the alternating spectrotemporal gaps. (B.) A colour-coded illustration of the relationship between the modulation patterns applied to the STG noise in the time domain (left) and frequency domain (right).

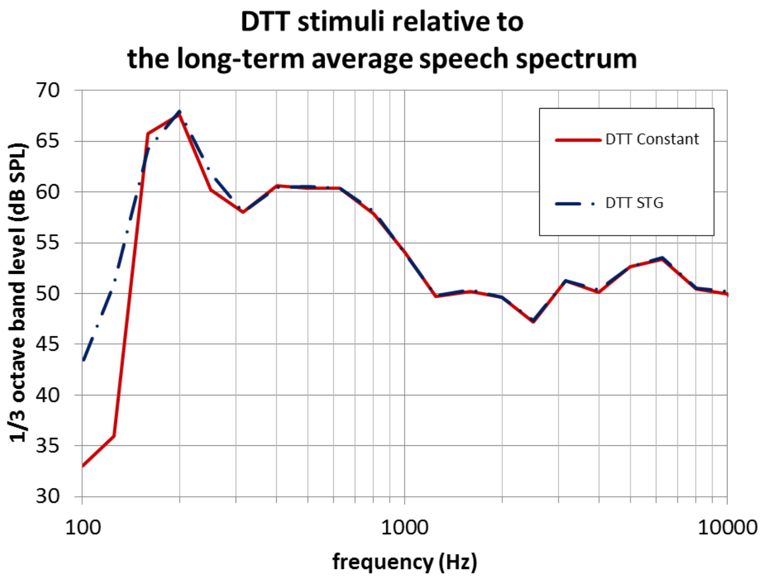


Figure 18: Comparison between the spectra of the test specific noise and the spectrotemporal gap noise.

3.4.2 Internet & telephone applications

An important aim for this study is to produce a digit triplet test that is applicable for telephone use as well as via the internet, specifically by using headphones. Over 67.5% of the Malaysian population has internet access (World Bank Group, 2014), hence it would be advantageous to use the infrastructure to access more users. Hearing screening via internet was also found to be the more favourable method of screening compared to telephone or visiting hearing care centres. It was also found to be the most cost effective method screening in terms of implementation and post-screening cost of intervention (Linssen et al., 2015).

The headphone and telephone applications necessitate separate normalisation in order to obtain an accurate equivalent lists for multiple applications together with their designated noise. The major difference between the two applications will be the frequency response of the transducers, where the telephone will have a narrower bandwidth compared to headphones. This has found to have an influence to speech perception performance of listeners where using telephones have been found to reduce SRTn in normal hearing listeners by up to 3 dB (Zokoll et al., 2012).

To normalise and evaluate the digits, the transducers were connected to an external sound card (Sound Blaster X-Fi Surround 5.1 Pro, Creative Labs, Singapore). For delivery via headphones, a Sennheiser HD 280 Pro headphone (Sennheiser electronic GmbH & Co., Germany) was used for all the testing throughout this study. The technical data for this model is as shown below.

Table 12: Technical data of the Sennheiser HD 280 Pro.

Frequency response	8-25000 Hz
Sound pressure level (SPL)	102 dB (IEC 268-7)
Total harmonic distortion	0.1 %
Nominal impedance	64 ohms
Load rating	500mW

For the telephone application, the handset (Cisco Unified Phone series 7900, Cisco Systems, Inc., CA, USA) was connected to a telephone tap (JK Audio THAT-2, JK Audio, Inc., IL, USA) which is then connected to the same external sound card. The technical data of the handset and audio tap is shown below.

Table 13: Technical data of the Cisco unified series 7900 handset.

Frequency response (narrowband)	300-3400 Hz
Microphone type	Electret
Impedance	8 ohms

Table 14: Technical data of the JK Audio THAT-2 audio handset tap.

Line inputs	RCA, 600-9000 ohms, 250mV RMS XLR female, 600-9000 ohms, 250mV RMS
Line outputs	RCA, 0-2500 ohms, 100 mV RMS XLR male, 0-2500 ohms, 100 mV RMS
Input gain	+12 dB maximum
Handset interface bias selection	Electret, dynamic and carbon

3.4.3 Software

Software was written by Assoc. Prof. Greg O'Beirne using the LabVIEW development platform (National Instruments, TX, USA). Specific modules were developed for the MDTT for the process of normalizing and evaluating the test.

3.5 Word and sentence structure selection for the Malay auditory-visual matrix sentence test.

3.5.1 Introduction

One of the important aspects of the matrix sentence design is the fixed syntactic structure that uses combinations of 50 predetermined words to form five word sentences that are highly redundant but unpredictable. Using this combination, a possible 100 000 sentences (5 word categories and 10 possible word each category = 10^5 sentences) can be derived from this structure. The published versions of the test in different languages are able to be compared, as they each use variations of the same fixed sentence structure with slight differences to suit the target languages (e.g. object following adjective for Malay but adjective follows object in English). Previous versions have taken into account the phonemic balance in the words chosen so that the test is more representative of the common phonemes used in the language. The reference materials between tests are not the same, as some languages may not have a standardized corpus of the most common phonemes. For example, the phoneme distribution in the Danish MST was compared to the 5000 most frequently used words, whereas the Turkish MST was not compared to any references as they assumed that the size of the base matrix and use of selected words were sufficiently representative of the language (Zokoll et al., 2015). The main objective of this part of the study was to determine the most appropriate sentence structure and words to be used in the MMST-AV.

Based on previous matrix sentence tests (see 2.6.3), it was decided that the Malay version would use the following sentence structure:

Noun + Verb + Number + Object + Adjective

3.5.2 Word selection for the Malay matrix sentence test.

In previous study by Mukari & Said (1991), words were selected from a Malay dictionary based on authors’ choice and were given to 150 Malay adults to rate in terms of familiarity. Lists were generated to ensure phonemic balance is achieved between lists but based on the phonetic transcription in Table 8, this aim was not achieved. The reason words were subjectively rated was because there is currently no corpus for Malay words and phonemes to draw this data from. Based on previous literature, if a corpus is not available for reference; one can be generated by comparing target stimuli with suitable references in languages printed documents, dialogues transcripts or other standardized speech tests (see Table 6: Comparison of language structure in selected international versions of matrix sentence tests.).

In order to select suitable words for the Malay MST, words were retrieved from the Institute of Language and Literature Malaysia website (www.dbp.gov.my). The website generated a corpus of Malay words depending on source of articles and its year of publication. A list of most frequent words was generated from Utusan Malaysia which is a mainstream Malay daily newspaper. Utusan Malaysia was chosen as the reference material as it is a commonly read mainstream newspaper covering all aspects of news including domestic and international affairs, politics, sports and entertainment. A total of 17,534 articles were analysed from the year 1980 up to 2012 and the analysis produced a compilation of 85,521 most frequent words.

Words for the categories subject, verb, number, adjective and object were selected according to their frequency, semantic neutrality and grammatical correctness. In this study, two syllable words were chosen since this is the most common type of word that can be found in Malay Language.

The inclusion and exclusion criteria for selected word are listed below.

Inclusion criteria:

i. Malay words
ii. Disyllabic words.
iii. Spondee
iv. Words that would fit into the five word categories listed above.
v. Words that were within top 2000 most frequent words found in the generated list , as they are most likely to be more familiar.

vi. Root words without affixes.

Exclusion criteria

i. Words that contain negative meaning or undertones.

ii. Special names to avoid bias. Words such as "*Anwar*" and "*Najib*" were excluded even though these words have among highest occurrence in the analysed newspaper. This is because these names belong to public figures in Malaysia with opposing political ideology.

iii. Words that contain a less-commonly occurring phoneme if they disrupt the phonemic balance.

Selected words were phonetically transcribed and compared to the phoneme distribution of the first 1000 most frequent Malay words in the daily. This is because the first one thousand most frequent words in a text corpus can represent as much as 72% of the source language use (Laufer & Nation, 1999). Selected words and their IPA transcriptions were then reviewed by a linguist who specializes in the Malay phonemes to ensure the syntax and semantic validity as well as the accuracy of transcriptions.

3.5.3 Results and analysis

The following words were found to be suitable in the matrix sentence test format, an English translation of the words are included below:

Table 15: Basic test list of the Malay matrix sentence test (top) and its English translation (bottom)

Subject	Verb	Number	Object	Adjective
Saya	bagi	satu	bola	baru
Kita	ada	dua	buku	besar
Dia	dapat	banyak	baju	lama
Kami	perlu	semua	lampu	kecil
Ibu	beri	tiga	meja	merah
Abang	ambil	empat	kotak	hitam
Ayah	mahu	lima	kunci	putih
Adik	suka	enam	pisau	hijau
Kakak	nampak	tujuh	mangkuk	mahal
Nenek	minta	lapan	topi	cantik

English translation

Subject	Verb	Number	Adjective	Object
I	give(s)	one	new	ball(s)
We	have/has	two	big	book(s)
He or She	receive(s)	many	old	shirt(s)
Us	need(s)	all	small	light(s)
Mom	give(s)	three	red	table(s)
Brother	take(s)	four	black	box(es)
Dad	want(s)	five	white	key(s)
Younger Sibling	like(s)	six	green	knife(s)
Sister	see(s)	seven	expensive	bowl(s)
Grandmother	asked for	eight	Pretty	hat(s)

The result of the comparison is shown below:

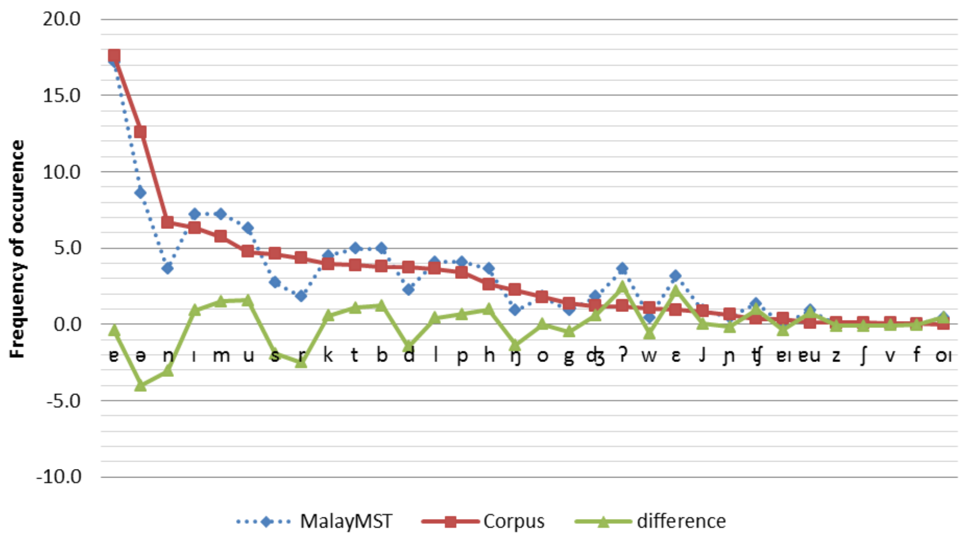


Figure 19: Comparison of phoneme distribution in percentage between selected words for the Malay matrix sentence test and the top 1000 most frequent words used in articles in Utusan Malaysia.

Table 16: Comparison between percentage of phoneme occurrence between words in the MMST-AV and the most frequent one thousand words in the Malay daily corpus.

Phoneme	Frequency (%)		Difference
	Selected Malay word list	Corpus	
m	7.24	5.26	-1.98
p	4.07	3.30	-0.77
b	4.98	3.48	-1.50
f	0	0.12	0.12
v	0	0.05	0.05
n	3.62	7.62	4.00
t	4.98	4.39	-0.59
d	2.26	3.16	0.90
s	2.71	4.59	1.87
z	0	0.14	0.14
r	1.81	4.14	2.33

l	4.07	3.57	-0.50
ʧ	1.36	0.40	-0.96
ʤ	1.81	1.35	-0.46
ʃ	0	0.14	0.14
ɲ	0.45	0.61	0.16
J	0.90	0.98	0.08
ŋ	0.90	2.33	1.43
k	4.52	4.33	-0.20
g	0.90	1.20	0.29
h	3.62	2.64	-0.98
ʔ	3.62	1.15	-2.47
w	0.45	1.30	0.85
ɪ	7.24	5.69	-1.55
ə	8.60	12.25	3.66
ɛ	3.17	1.10	-2.06
ɐ	17.19	17.76	0.566
u	6.33	4.65	-1.69
o	1.81	2.04	0.23
ɐ ɪ	0	0.20	0.20
ɐ u	0.90	0.06	-0.84
oɪ	0.45	0	-0.45
Absolute average value (∑ difference)			0.00

The association between proposed Malay words for the MMST-AV with the corpus was analysed and it showed a significant and strong positive phonemic agreement between these two word groups (Pearson's $r = 0.919$, $p < 0.001$). The average absolute difference in phonemic occurrence between the corpus and the proposed words for the MMST-AV is 0.00. Transcription of the most frequent 1000 words is attached as Appendix B.

3.5.4 Discussion

The words selected for the MMST-AV have close to equal phoneme distribution to the written corpus, with the largest difference observed in syllable /ə/ at 3.66% (around 30% less common in the test than in the corpus). There were many special names within the top 2000 most frequent words; however it was decided to keep generic names like *ibu* (mother) or *nenek* (grandmother) to ensure the test is neutral semantically. The possible word combinations are shown below:

Using Pearson's linear correlation, the words for the MMST-AV were found to correlate better to its reference corpus compared to the New Zealand version of the matrix sentence test in terms of phonetical balance (Trounson & O'Beirne, 2012) as shown below;

Table 17: Comparison on statistical finding between New Zealand English matrix with Malay-PWL

Matrix sentences	**Statistical finding**
New Zealand English matrix sentence test	$r = 0.6$, $n = 42$, $p < 0.001$
MMST-AV	$r = 0.919$, $n = 32$, $p < 0.001$

The differences are likely due to the source of speech material used in selecting the base word list. In the MMST-AV, words were selected and compared to the most frequent words from a mainstream newspaper from the year 1980 until 2012. In contrast with New Zealand English Matrix, the word list was adapted from British English word matrix and was compared to the phonemic content of the New Zealand hearing in noise test (NZHINT). The NZHINT test (Hope, 2010) was based on sentences compiled from children's book. The words selected for the MMST-AV were not compared to the phonemes of the Malay HINT (Quar et al., 2008) since it did not make any claim to represent the overall phonemic distribution for the adult population in the Malay language.

The limiting factor of using written materials as a reference is that it may only represent 72% of the words used in a complete corpus of language (Laufer & Nation, 1999). A corpus of spoken words may contain informal words that are not used a newspaper and can increase coverage of language to 84%. Further analysis of spoken languages in current movies or language samples together with the analysis of written words may add more value to the comparison.

3.6 Development of the Malay auditory-visual matrix sentence test

3.6.1 Introduction

To ensure the SRT and slope of intelligibility will be comparable to other established matrix sentence tests, the following methods were adapted from the work by Akeroyd et al. (2015), Trounson (2012) and Wagener et al. (2003).

3.6.2 Recording of speech and visual materials

The same speaker from the Malay DTT was recruited for this test. Video was captured using a Sony PMW-EX3 high definition video camera together with a condenser microphone (Model C568 EB, AKG, Vienna, Austria). The speaker was seated against a wall in an audiometric cabin and her head was supported by a head brace to maintain a stable head position throughout the recording (see Figures 23 to 26 for the recording set up in the audiometric booth). The video was recorded in high definition at 720p and 50 frames per second. The audio recordings were done at a sampling rate of 48,000 samples per second in the Pulse Code Modulated (PCM) format at 16-bit resolution. These parameters (particularly the 50 fps frame rate) were chosen to enable more precise editing and to allow for future manipulation of display resolution for larger displays. To ensure efficient smooth recording of sentences, the speaker was asked to read the sentences from an autocue system (see Figure 24) that projected sentences containing of all possible word combinations. The list of 100 sentences was designed to contain all 400 possible word pairs (see APPENDIX A). This method was suggested as it preserves natural co-articulation between words in the sentence (Wagener et al., 2003). Figure 22 indicates how all possible 400 word pairs were pooled.

Trounson (2012) reported that it was necessary to provide mechanical support for the head and neck as small movements during recording caused significant shifts between frames when the edited video where combined. The custom head brace was made from plaster gauze to help minimize head and neck movements. This head brace was covered during the recording session using the speaker’s hijab or headscarf which is a Muslim attire used widely around the world.

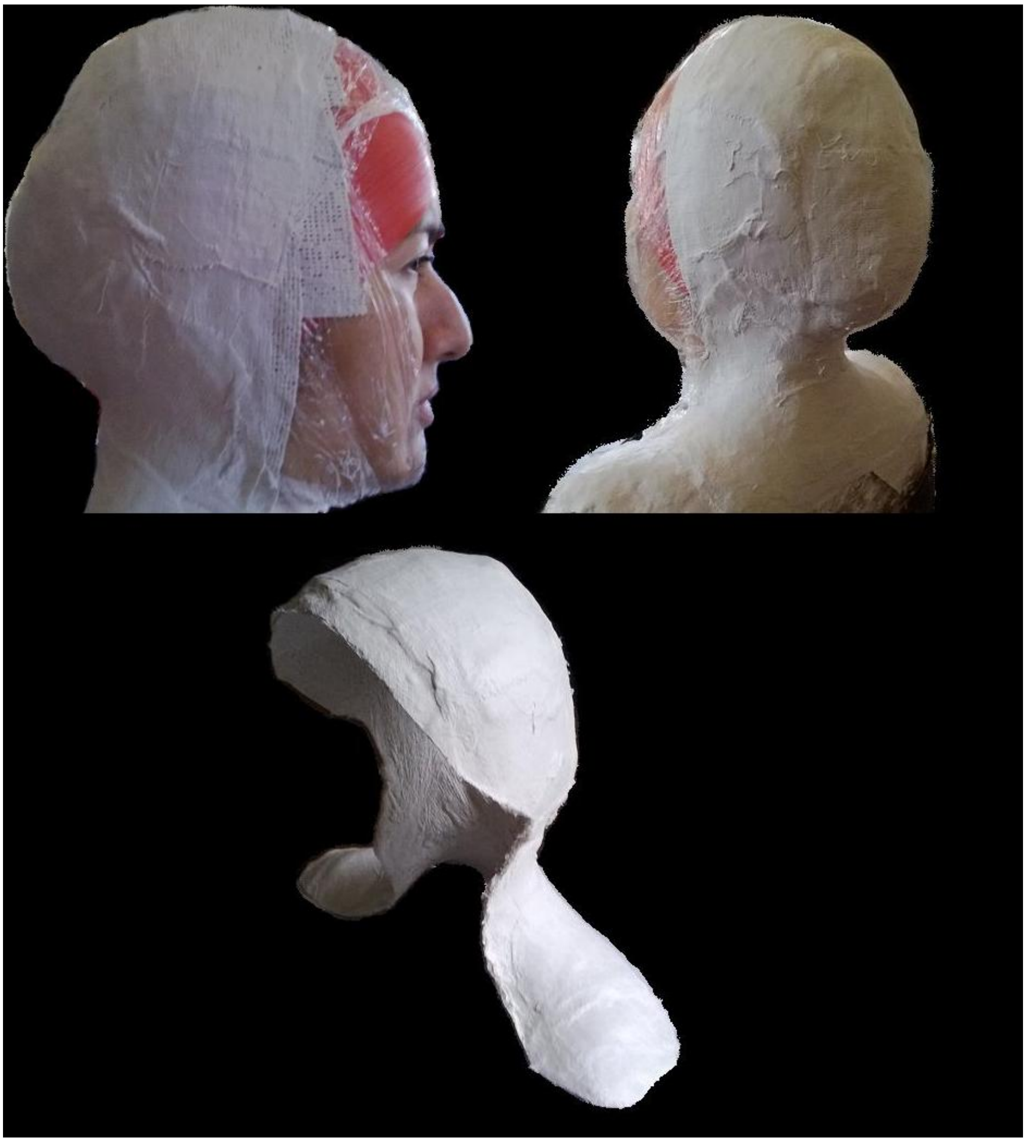


Figure 20: Custom head brace. Speaker was lined with food wrap plastic and swimming cap before plaster gauze was placed around the head, neck and shoulders. Once the plaster was set, the head cast was cut to an appropriate shape to make it easier for the speaker to put in on and off.

Saya	bagi	satu	bola	baru
Kita	ada	dua	buku	besar
Dia	dapat	banyak	baju	lama
Kami	perlu	semua	lampu	kecil
Ibu	beri	tiga	meja	merah
Abang	ambil	empat	kotak	hitam
Ayah	mahu	lima	kunci	putih
Adik	suka	enam	pisau	hijau
Kakak	nampak	tujuh	mangkuk	mahal
Nenek	minta	lapan	topi	cantik

Figure 21: Identifying all possible word combinations of the Malay matrix test.

Figure 22: Speaker’s head position during recording. A polystyrene box was cut into the shape of the head brace to hold the speaker’s head into place and she was also strapped to the chair at the shoulder and waist to secure body movement. The green-screen cloth was placed between the polystrene and the speaker during recording.

Figure 23: Autocue setup using cardboard box, a 45 degree angled glass and a smartphone to project sentences to the mirror.

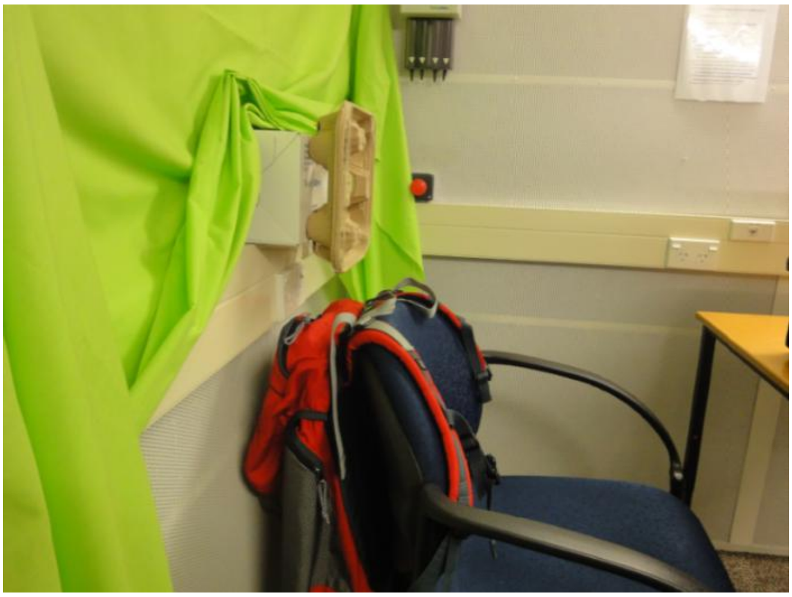


Figure 24: Recording chair set up. A large backpack was used to securely strap the speaker at the shoulder and waist.

Figure 25: Speaker seated at the recording position in the audiometric booth.

3.6.3 Editing of video and audio materials

3.6.3.1 Adjusting video output

The recorded video and audio file was edited using Adobe Premiere Pro CC software. Before the video was edited to obtain all 400 possible word pairs, the video files were enhanced to improve viewing experience. The steps taken are described in the figure below.

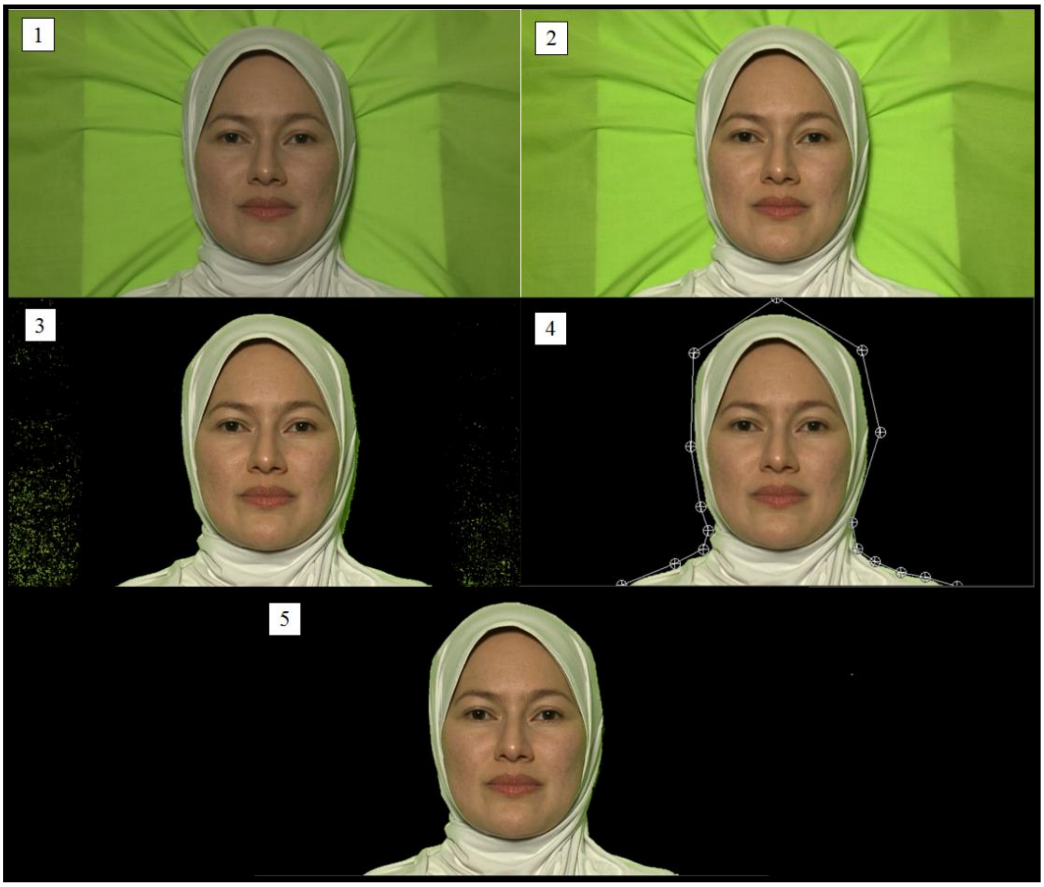


Figure 26: Video enhancements. Picture 1 shows the raw footage before any enhancements were done. Picture 2 shows adjustments of brightness and contrast. Chroma key was added as shown in picture 3, minimal application of chroma key distorted color saturation and picture clarity hence some green coloration is still seen at the vertical edges of the picture and close to the speaker's hijab. A 16-point garbage matte application was used to remove the green discoloration as seen in picture 4. Picture 5 shows final image used in the MMST-AV.

3.6.3.2 Sentence segmentation and word edit

The editing process includes making sure that natural transition between words are preserved both in the video and audio files. An overlap of 15 milliseconds between 'in' and 'out' points was not used based on the suggestion by Hochmuth et al. (2012). This is because video transition quality was affected when this was applied. Raw video footage was edited into smaller files to make it easier for further editing and labelling files. The raw video that contained all 100 sentences read by the speaker were broken down into 100 separate media files containing a sentence each. A critical part of the video and audio edit was the determination of the ‘in’ and ‘out’ points where important visual and audio clues can be lost and cause confusion during testing. Using the separate 100 files, media clips were segmented into suitable fragments. Observation of the Malay language structure suggested that the edit points used in Trounson (2012) may not be applicable for the Malay language. Therefore a new set of editing rules based on the Malay language was used.

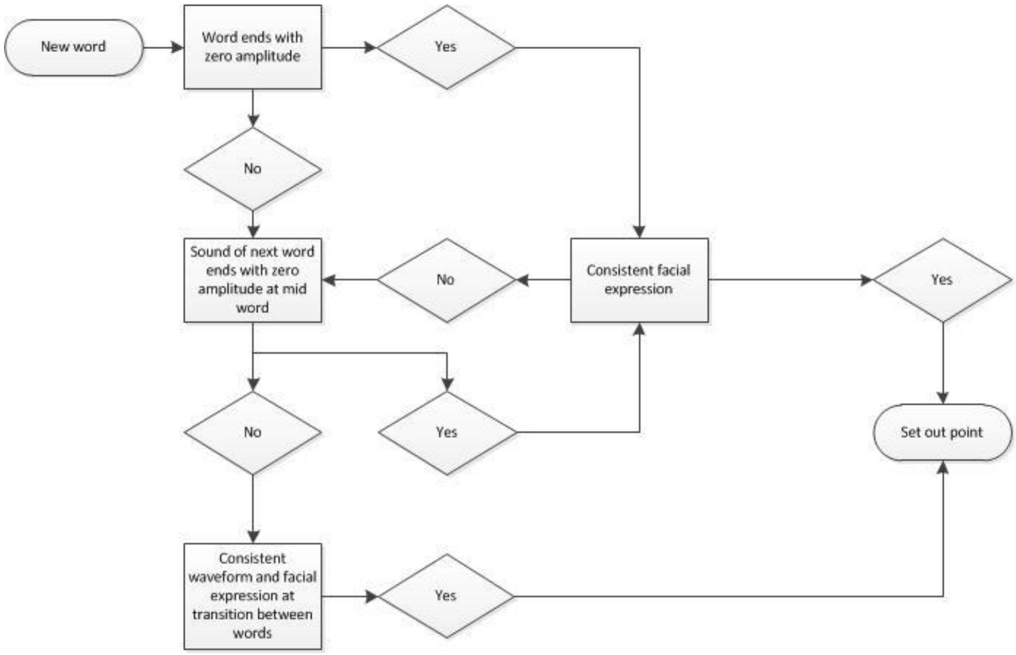


Figure 27: Workflow to obtain best ‘in and out’ points for the Malay matrix sentence test

An example of an edit is shown below:

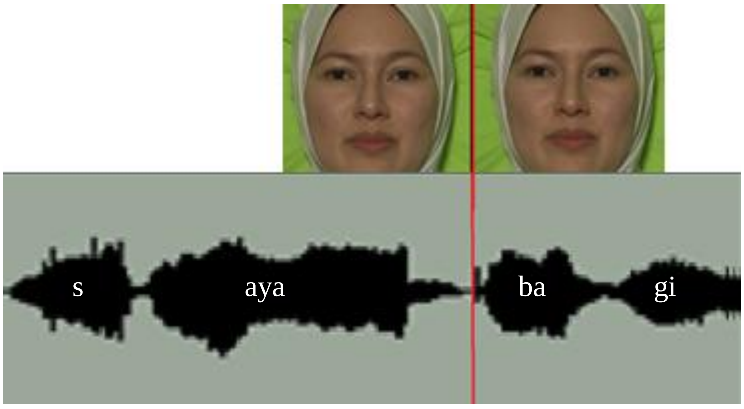


Figure 28: Example of between word edit of the words “saya bagi” or I give. The red line indicates the ‘out’ point for the word “saya” for this specific word combination and the ‘in’ point for the word “bagi”. Picture shows speaker’s mouth position before the noise burst for the phoneme /b/ was produced. All the edit points were counter examined between visual and auditory spectral changes to ensure consistency between frames.

Table 18: Table showing all possible edit cuts for the MMST-AV. The “||” symbol indicates the position of cuts in between or within words.

<table>
<tr><th>Subject</th><th>Verb</th><th>Number</th><th>Object</th><th>Adjective</th></tr>
<tr><td>Saya</td><td>||bagi</td><td>s||atu</td><td>||bola</td><td>baru</td></tr>
<tr><td>Kita</td><td>a||da</td><td>||dua</td><td>||buku</td><td>besar</td></tr>
<tr><td>Dia</td><td>||dapat</td><td>||banyak</td><td>||baju</td><td>lama</td></tr>
<tr><td>Kami</td><td>||perlu</td><td>s||emua</td><td>||lampu</td><td>kecil</td></tr>
<tr><td>Ibu</td><td>||beri</td><td>t||iga</td><td>||meja</td><td>merah</td></tr>
<tr><td>Abang</td><td>||ambil</td><td>||empat</td><td>||kotak</td><td>hitam</td></tr>
<tr><td>Ayah</td><td>||mahu</td><td>l||ima</td><td>||kunci</td><td>putih</td></tr>
<tr><td>Adik</td><td>s||uka</td><td>||enam</td><td>||pisau</td><td>hijau</td></tr>
<tr><td>Kakak</td><td>||nampak</td><td>||tujuh</td><td>||mangkuk</td><td>mahal</td></tr>
<tr><td>Nenek</td><td>||minta</td><td>l||apan</td><td>||topi</td><td>cantik</td></tr>
</table>

3.6.3.3 Encoding media

A re-encoding process was necessary to change the original recording format to a file format compatible with a standard version of Windows Media Player which was used in the UCAST software to play media files. The other reason for this is because almost all personal computers running Microsoft Windows are supplied with this media player, making it easy to download UCAST application to many PCs. The audio and video files were encoded separately using a freeware called *FFmpeg* (www.ffmpeg.org). The final versions of audio and video media files were labelled as the names of word pairs it represents for example, *ayah_mahu.mpg* and *ayah_mahu.wav* or *tiga_pisau.mpg* and *tiga_pisau.wav*.

Table 19: Video and audio format of original footage and final MMST-AV audio and video file encodes.

	Original recording	Fragment storage format	Final MMST-AV stimulus
Video format	.mp4	.mpg	.avi
Video codec	MPEG-1/2 video (mpgv)	MPEG-1/2 video (mpgv)	MS MPEG-4 video v3 (MP43)
Video frame	1280 x 720	936 x 702	640 x 480
Color depth	32 bit	32 bit	32 bit
Aspect ratio	16:9	4:3	4:3
Fields	Progressive	Progressive	Progressive
Audio format	.wav	.wav	.wav
Audio channel	Stereo	Mono	Mono
Audio codec	PCM s16 BE (twos)	PCM S16 LE (s16l)	PCM s16 LE (araw)
Audio sample rate	48 kHz	44.1 kHz	44.1 kHz
Audio bit depth	16 bit	16 bit	16 bit

The word pair media files will need to be concatenated to form a complete sentence. Using .mpg formats, it was found that the transitions between video frames are smoother hence *FFmpeg* was used to convert the media files in the following order.

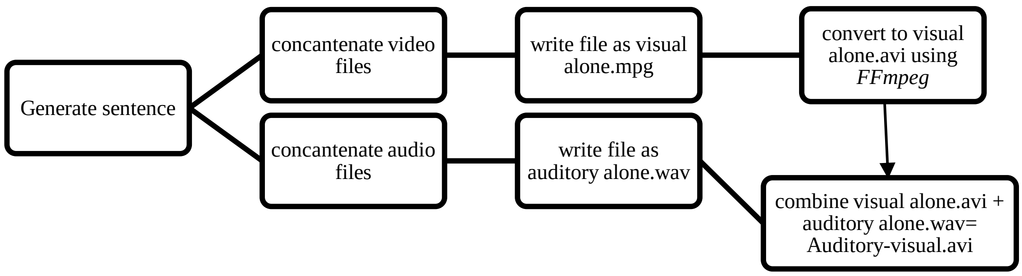


Figure 29: File encoding sequence

3.6.4 Software development

Software for user interface was developed by Assoc. Prof Greg O'Beirne using National Instruments Labview development environment (version 12). Audio and video files of words were designed to be pooled together in real time as this was seen to be an efficient method of producing media samples without taking large amounts of storage for media files (refer to Figure 30: File encoding sequence). The interface was designed to present the MMST-AV in auditory alone, visual alone or auditory-visual modes. For the normalisation process, the software or virtual instrument (VI) was designed to present all the stimuli at fixed SNRs. Separate VIs using the same LabVIEW platform was designed specifically to evaluate the audio recordings at fixed SNRs and also the visual quality of the test. Finally the MMST-AV was incorporated into the University of Canterbury adaptive speech test (UCAST) software interface for the purposes of validation of the test. Several versions of each module were updated to overcome certain programming issues in the software.

In general all the versions of the software perform almost similar sequences to present and record measurements. The sequence of operations is shown in the figure below.

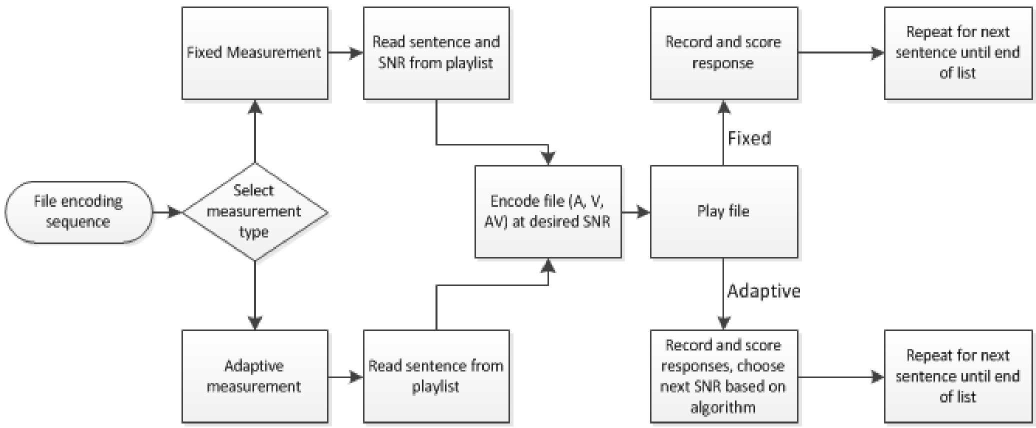


Figure 30: Code sequence for the MMST-AV software.

The interface for the software to normalise and evaluate the MMST-AV used a close response method by displaying the word matrix. Listeners were required to click on the word of choice or point to the word if a touch screen monitor was available. For the auditory-visual mode, the screen would only display the speaker uttering the words on a black background and the word selection would only appear immediately after the speaker has completed the sentence. For the auditory-alone mode, the screen was similarly black during stimulus playback. This was to ensure listeners were focused on the listening and lipreading tasks during the test instead of browsing through the word options.

3.6.5 Development of masking noises for the Malay matrix sentence test

3.6.5.1 Test specific noise or test specific noise (TSN)

The test specific noise was designed to provide energetic masking (Brungart, Simpson, Ericson, & Scott, 2001) and is also the optimum masker for the matrix sentence tests (Kollmeier et al., 2015). Using similar techniques used in the Malay DTT, a speech shaped masking noise was generated by superimposing the recorded word pairs 10000 times within a 10 second looped sound file using an automated process.

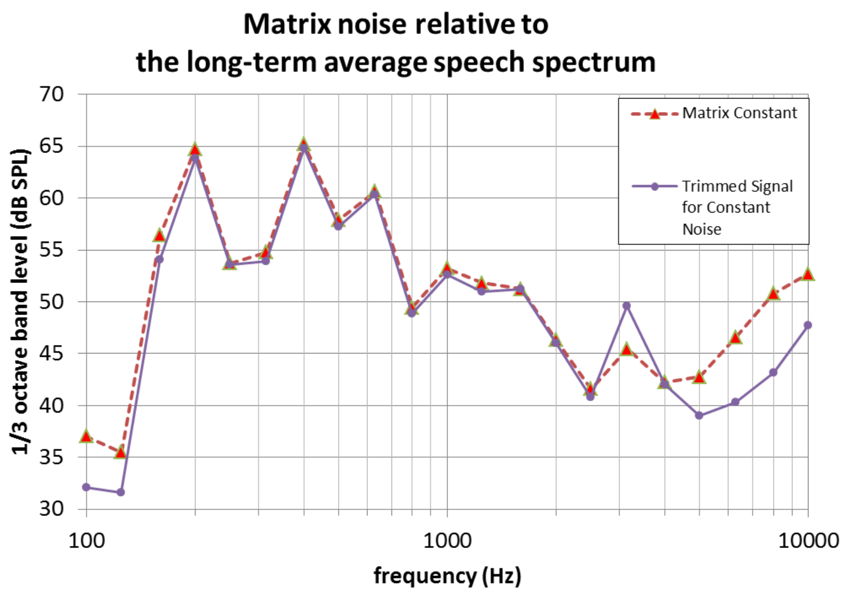


Figure 31: Spectrum of signal and speech-shaped noise used in the MMST-AV.

3.6.5.2 Malay 6-talker babble noise (BN)

As the Malay matrix test was designed to test “real-world” speech perception abilities, a constant noise may not be a common element in masking speech sounds in real life scenarios. Therefore to further enhance face validity of this test, a custom Malay babble noise was created. This masker was designed to provide informational masking specifically for the MMST-AV (Freyman, Balakrishnan, & Helfer, 2004). Three male (including author) and three female native Malay adult speakers volunteered to provide their voices for this noise. The volunteers were asked to read a monologue from a newspaper article titled “Cahaya takdir” (retrieved from www.utusan.com.my dated 7th December 2013). The monologue was used to replicate conversations in a cocktail party like atmosphere with six talkers as background noise. The monologue used normal everyday words. The audio was recorded using the TEAC TASCAM (model HD-P2 Portable High Definition Stereo Recorder) in the same audiometric cabin used by the main speaker of the Malay matrix test. Each of the audio files was then edited to 5 second audio loops which were then shuffled using an automated process.

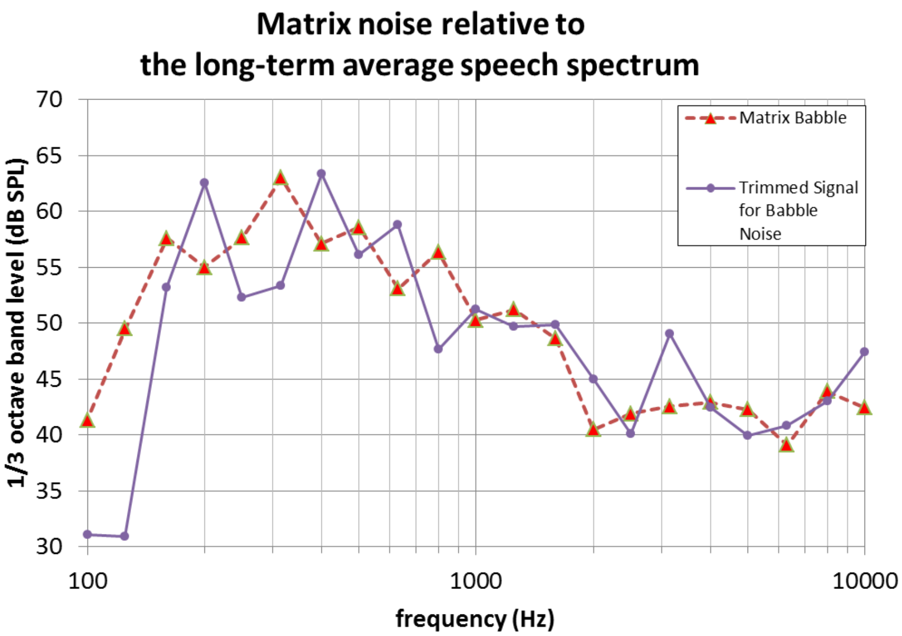


Figure 32: Frequency spectrum of signal and 6-talker babble noise used in MMST-AV.

3.6.6 Verifying video edit transitions

To verify whether the video edits could result in smooth visual transitions between the last frame and the first frame in a sentence, Assoc. Prof Greg O'Beirne wrote a VI to calculate pixel differences between frames for all of the 400 word pairs to a reference video frame. A template or reference video frame was taken from the last frame form the video of the word pair "Saya bagi". Parts of the mouth, nose and jaw were removed from this comparison, as these parts of the face move the most in normal speech. Subsequently 10 pixels were subtracted from the top, bottom, left and right from each video frame using *FFmpeg*, and the cropped image was then shifted both horizontally and vertically over the reference template to determine which overlay location produced the smallest difference in pixel values. In most cases, the location corresponding to middle and centre of both image overlying exactly produced the smallest difference value, indicating that the images were already in good alignment. However, our investigation showed some significant shifts in 3 particular word pairs, as shown in Figure 33.

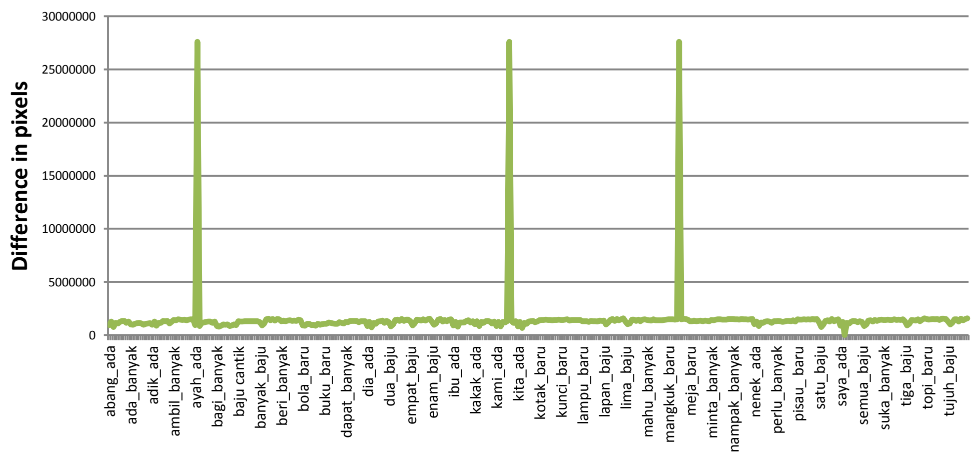


Figure 33: Pixel difference between video segments and reference template.

Inspection of transition between words revealed big jumps between frames. After close inspection of the video recording, there were changes in frame in the vertical axis when words recorded from early parts of the video were concatenated to video of words from the end part of the video. This is possible due to movement of the camera person (in this case the author during recording) as the floor of the audiometric cabin flexed when the camera person changed gait in the room. Instead of removing the video segments from the test, another VI was written to make the appropriate shifts to the nearest pixel to all video frames using the same reference template. The VI calculated the least amount of pixel difference between all video files and the same reference template. The result of this adjustment was deemed satisfactory based on subjective observation of videos of sentences manually concatenated together. To evaluate the effectiveness of the editing process and this adjustment, a group of normal hearing native Malay speakers evaluated video quality of the MST-AV, and these results are discussed in Chapter 4.

3.6.7 Scoring the MMST-AV

Due to the unique way the words were edited, some media files contained parts of other words. This has led to a necessary revision of the word scoring methods commonly used in matrix sentence tests. Intelligibility functions were generated for both fragment and words where all selected words were categorized into two parts to map the media fragments equally. This method also allowed scores to be measured using the actual fragments that were used to present the words. Example of the scoring method is shown below. In the table below, the fragment "Ibu_suka" was perceived as "Ibu mahu". For the word scoring method, this is interpreted as a score of 1/2 or 0.5 as only the word Ibu was correctly perceived. However in the fragment scoring method, two out of three halves were scored correctly, giving a score for that fragment of 2/3 or 0.667.

Actual	Ibu		suka		lima		buku		cantik		Fragment scoring		
Selected	Ibu		suka		lima		buku		cantik		**Pt 1**	**Pt 2**	**total**
Ibu_suka	I	bu	s								2/2	1/1	1
suka_lima				uka	l						1/1	1/1	1
lima_buku						ima					1/1	0/0	1
buku_cantik							bu	ku	can	tik	2/2	2/2	1
Word scoring	1		1		1		1		1				

Actual	Ibu		suka		lima		buku		cantik		Fragment scoring		
Selected	Ibu		suka		lima		baju		cantik		**Pt 1**	**Pt 2**	**total**
Ibu_suka	I	bu	s								2/2	1/1	1
suka_lima				uka	l						1/1	1/1	1
lima_buku						ima					1/1	0/0	1
buku_cantik							bu	ku	can	tik	0/2	2/2	0.5
Word scoring	1		1		1		0		1				

Actual	Ibu		suka		lima		buku		cantik		Fragment scoring		
Selected	Ibu		mahu		empat		buku		cantik		**Pt 1**	**Pt 2**	**total**
Ibu_suka	I	bu	s								2/2	0/1	0.667
suka_lima				uka	l						0/1	0/1	0
lima_buku						ima					0/1	0/0	0
buku_cantik							bu	ku	can	tik	2/2	2/2	1
Word scoring	1		0		0		1		1				

Figure 34: Scoring procedure for the matrix sentence using the example sentence of *'ibu suka lima buku cantik'* or "mother likes 5 beautiful books".

3.7 Calibration

Both tests use signal-to-noise levels as means to identify speech perception in noise, hence it is relatively free from issues of using different types of transducers and sound processing apparatus (Smits et al., 2004). To ensure the signal-to-noise ratio levels were at the determined level, the test setup was calibrated using the GRAS ISO 4869-3 Hearing Protector Test Fixture Type 45CA (1 second averaging time) connected to the Brüel & Kjær type 3560 C pre-amplifier. Sound level monitoring was recorded using the Brüel & Kjær – PULSE Labshop fast track version 17.1.1 software. The Sennheiser HD 280 headphone was coupled to the GRAS artificial ear while running the UCAST platform with stimuli and noise level was set separately at a constant 65 dB A for calibration purposes. Before measurements could be done all the silent gaps between speech sounds were removed as it affects the overall equivalent power level of the sound. Measurement averaging time was set at 32 seconds. All calibration figures were noted and were used to adjust to all SNRs measured in both tests.

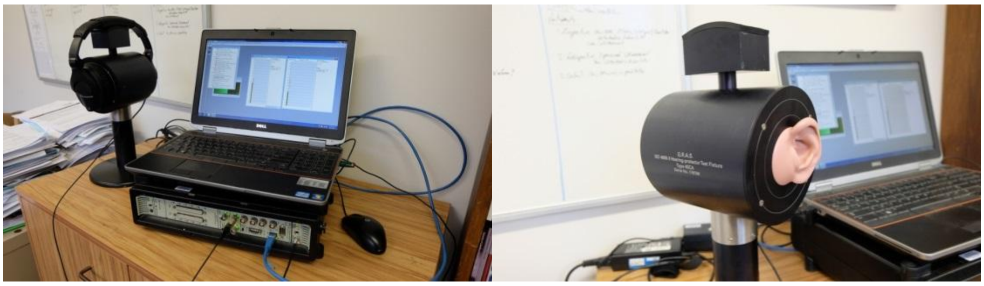


Figure 35: Calibration setup

3.8 Discussion

In this chapter, the development of the MDTT and MMST-AV was described. Special consideration was given to ensure the quality of the recordings was good, and included several troubleshooting runs before the recording sessions to identify possible problems which may have arisen (not described here).

3.8.1.1 Speaker selection

The speaker was selected based on overall score of subjective ratings from adult Malay native speakers. Both speakers were rated equally for the voice quality but speaker 1 was rated easier to lip read. Evidence shows that lip reading is an important component to speech perception, as infants have been shown to have language-specific lip reading skills as early as 4 months (Woodhouse, Hickson, & Dodd, 2009). Lip reading improves speech perception dramatically, especially in difficult listening condition. It also known that lip reading ability reduces with age and not gender specific (Tye-Murray et al., 2007c). Taking this into consideration, a speaker that is easier to lip read will be more practical as the MMST-AV was designed for a broad spectrum of adult listeners across a range of ages and hearing abilities.

3.8.1.2 Development & recording of MDTT

Digit selection for the MDTT was mostly a direct process, as no issues relating to digit homogeneity were encountered after eliminating the tri-syllabic digit 9. Two takes of digit recordings were used for the purpose of normalisation. This allowed the option to choose the best digit recordings between the two takes using the slope of intelligibility of each digit.

Understanding that there is room for improvement in test sensitivity, especially for sloping hearing losses, we developed the STG noise. The modification to the speech-shaped noise was to encourage release from masking for normal hearing listeners. The results obtained using these tests are described in Chapters 4, 6 & 7.

3.8.1.3 Development and recording of MMST-AV

The Malay sentence structure selected for this MMST-AV is similar to the French, Italian and Spanish MSTs (Hochmuth et al., 2012; Jansen et al., 2012; Puglisi et al., 2015). Hence, the language structure is expected to have no influence for this test as it has retained the same “5 homogenous words per sentence” structure (Hochmuth, Jürgens, et al., 2015).

The Malay words selected seem to represent well the overall phonemic distribution of the reference corpus. Future evaluation of the phoneme distribution can use other sources of informal language samples or standardized Malay tests such as the Malay HINT. This is however not the most important aspect of the test content, as there are other aspects of speech and language that could affect test results, such as temporal cues of speech and inter talker differences (Hochmuth, Jürgens, et al., 2015).

Recording the MMST-AV presented a challenge, as both video and audio recording integrity had to be ensured. Using recommendations by Trounson (2012), the quality of the video recording was improved immensely. It is recommended that for future recordings, the speaker and camera are either situated on a concrete floor, or the speaker is left alone to complete the sentences. It was convenient for this study to use a plaster head cast as it was hidden underneath the speaker’s hijab. The head cast proved to be an important key to good recording, however, the plaster cast may be too thick and visible during recordings of speakers who do not wear the hijab. We therefore propose for future recordings that a 3D scan be made of the speaker which can later be used to design a head cast mesh seen below. The material used in 3D printing is similar to the plastic used to make LEGO blocks and it can be made as thin as 2 millimetres, so the cast should be more durable than the plaster gauze. The mesh could be easily covered superimposing an image or with some creative hair styling.

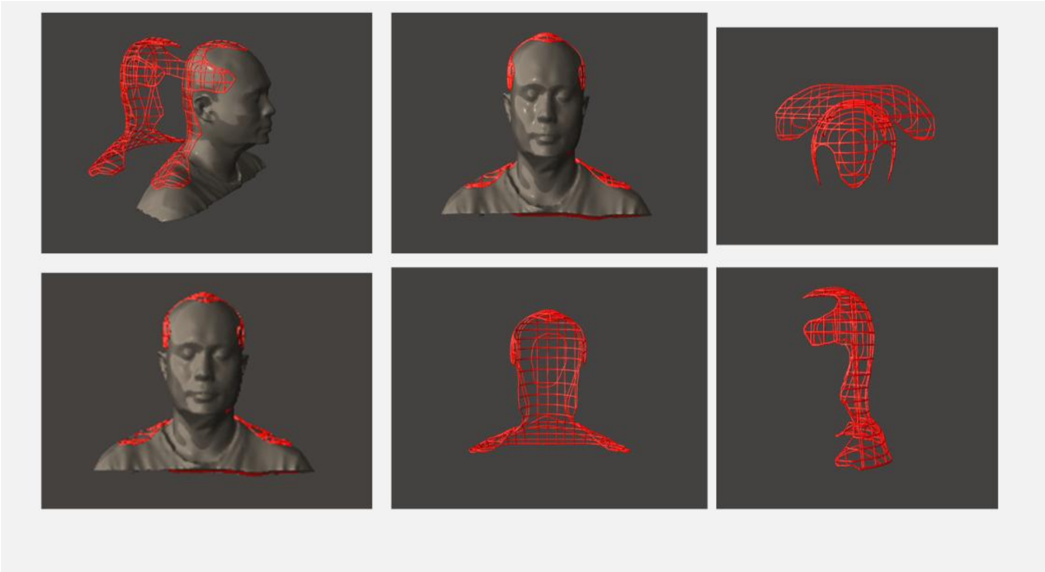


Figure 36: 3D scan of the author and a custom head mesh that can be printed.

CHAPTER 4

NORMALISATION OF THE MALAY DIGIT TRIPLET AND MATRIX SENTENCE TESTS

4.1 Introduction

Normalisation (also called optimization) is a crucial step in the process of constructing auditory stimuli that could help make the test more accurate and reliable. This process is done through estimation of a certain point of the intelligibility curve (usually at 50% intelligibility) using a logistical function. The direct benefit of this procedure is seen in the improvement of intelligibility slopes after the level adjusting the stimuli to match the average intelligibility score for all stimuli. The basis of this process can be explained using the probability function below.

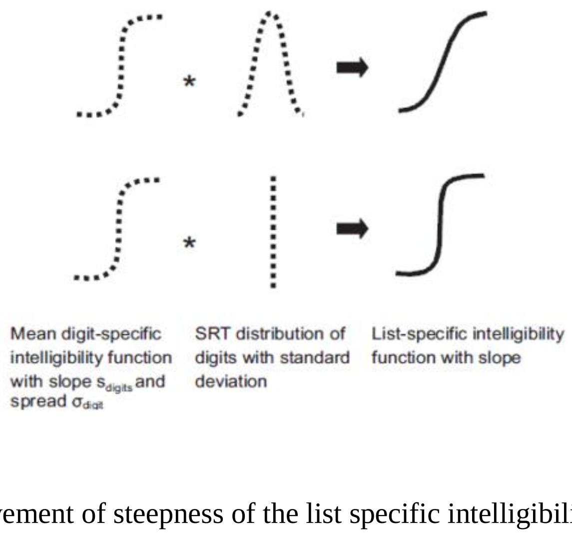


Figure 37: The improvement of steepness of the list specific intelligibility (S_{List}) functions is due to the smaller digit-specific standard deviations as the result of level adjustments made to digit-specific SRT values (Adapted from Zokoll et al., 2012).

Equation 1: The relationship between digit-specific SRT and list-specific steepness as explained by the probabilistic model by Kollmeier (1990) as cited by Zokoll et al., (2012).

Despite using different approaches to normalise their materials, the German and Polish DTT studies reported very similar average slope of intelligibility. This was achievable as both tests used psychometric procedures to estimate the SRTn, which is the basis of the level adjustments. Both DTTs and MSTs have shown considerably higher intelligibility slopes than other speech tests in noise, for example the HINT test (Soli & Wong, 2008). The French DTT test showed the steepest slope of intelligibility at 27.1%/dB, whereas the shallowest slope for a DTT was reported for the Dutch version at 16%/dB (Zokoll et al., 2012). For the MST, the steepest slope was recorded by the Polish MST at 17.15 %/dB, and the shallowest slope in an MST test was found in the Dutch version at 10.2%/dB. The improvement of list-specific slopes directly translates to the ability of the test material to differentiate speech perception ability in noise for listeners. In comparison between the two tests, the DTT showed higher average slope of intelligibility compared to the MST and is consistent with previous findings (Miller et al., 1950). The MST uses sentences that carry more speech and language context compared to digits, hence it would be more difficult to guess digit combinations than word combinations in the MST.

This chapter describes the normalisation performed for the Malay versions of the DTT. The normalisation of the MDTT was done using headphones and telephones in two types of background noise, which are the test specific noise or test-specific noise (TSN) and the spectrotemporal gap noise (STG).

4.2 Normalisation of Malay digit triplet test

The aim of this study was to normalise digit triplets in Malay using normal hearing listeners. The normalised digits were to be used to form equally intelligible lists.

4.2.1 Method

4.2.1.1 Sample size calculation for the Malay DTT

In order to study repeated measures of the DTT for normalisation purposes, the sample size of the each group was calculated using a power and sample size calculator software (Dupont & Plummer, 1990). The target power (the ability to reject the null hypothesis in favour of a specific true alternative) was set to a probability of 0.9 - the Type I error probability associated with this test of this null hypothesis was 0.05. In a previous study (Zokoll et al., 2012), the response within the normal hearing group was normally distributed with standard deviation of less than 1 dB, which indicated that at least 13 normal hearing (HTL 0.5,1,2,4 kHz < 20 dB HL) subjects for each normalisation step should be recruited. Based on previous work by Murray (2012) and King (2011) for the Māori and New Zealand digit triplet tests, a larger number of samples for the normalisation process would be preferable for a more precise fit of the psychometric curve.

4.2.1.2 Subject recruitment

Twenty normal hearing (average hearing threshold levels at 250, 500, 1000, 2000, 4000 & 8000 Hz ≤ 20 dB HL) Malay native speakers were recruited at this stage of the study using convenience sampling. All participants were aged 18 to 43 years (average, 31.7 years; S.D., ± 9.2 years) and had no history of ear and balance problems. Participants were either students or staff at the Department of Audiology and Speech-Language Pathology at the International Islamic University Malaysia, Kuantan Campus. Participants were paid for their involvement.

4.2.1.3 Testing procedure

All participants were tested at the IIUM Hearing & Speech Clinic, in Kuantan Malaysia. They were asked questions about their history of ear health and tested for hearing threshold levels at levels at 250, 500, 1000, 2000, 4000 & 8000 Hz. The digit triplet test was presented using a custom VI created by Assoc. Prof. Greg O’Beirne (user interface shown in Figure 38 below). The VI was running on a Windows PC using an external sound card (Sound Blaster X-Fi

Surround 5.1 Pro, Creative Labs, Singapore). For the headphone and telephone normalisation the transducers Sennheiser HD 280 Pro headphones (Sennheiser electronic GmbH & Co., Germany) and Cisco Unified Series 7900 telephone (Cisco Unified Phone series 7900, Cisco Systems, Inc., CA, USA) were used. The telephone handset was directly coupled to the sound card via JK Audio THAT-2 audio handset tap (JK Audio THAT-2, JK Audio, Inc., IL, USA). For both headphone and telephone normalisation, participants were tested at five different SNRs, which were -22, -18.5, -15, -11.5 and -8 dB SNRs using the list of 180 triplets. Background noise was presented monaurally at a steady 65 dB SPL throughout testing and the signal-to-noise ratio were adjusted by changing intensity levels of digit triplets. In total all participants were presented with 720 triplets (180 triplets x 2 masking noise types x 2 transducer filter settings x 2 presentations) for all four test conditions:

i. Headphone in TSN;

ii. Headphone in STG;

iii. Telephone in TSN;

iv. Telephone in STG.

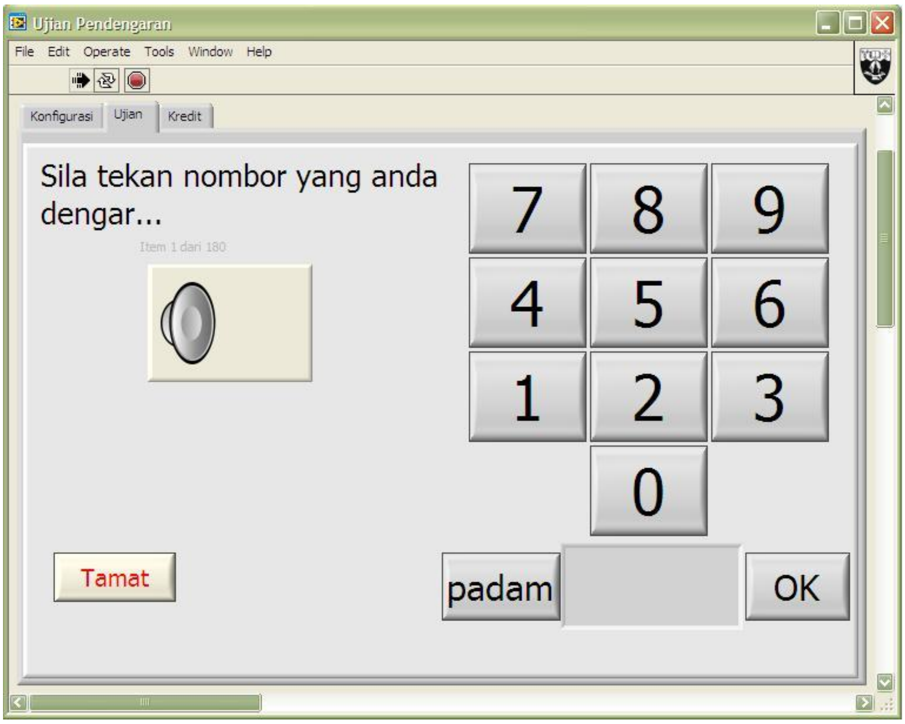


Figure 38: Graphical user interface used for the normalisation process. The text reads "Please press the numbers you hear".

All participants were seated in a double walled audiometric booth (average ambient noise = 22.5 dB SPL, reverberation time R_{T60}= 0.1 millisecond) and were asked to key in their responses on a keyboard. A standard written instruction was provided in the interface before test started. Participants were asked to guess if they were unsure of the digits they heard – however, to reduce the effect of chance score for each digit tested, all digits were presented twice for each condition. Data from all the participants were pooled to enable the calculation of psychometric functions for each digit recording in each of the positions – for example, for the triplet "123", "1" is in the first position, "2" is in the second position, and "3" is in the third position. The digit recordings from either Take 1 or Take 2 that generated curves with steeper slopes than their alternative recordings were used for the evaluation process of this test.

4.2.1.4 Determining SRT levels

The SRT for all digits were level matched as closely as possible to the mean SRT for all digits at each position by adjusting the signal levels by not more than 4 dB. The following logistic function was used to determine the intelligibility level for each digit:

Equation 2: Digit-specific logistical function as described in Zokoll et al., (2012), with *y* being chance score and s for slope at the SRT.

4.2.2 Results and analysis

For each digit at each digit position for both Take 1 and Take 2, the intelligibility function shown in Equation 2 above was fit to the data (the mean percentage correct at each of the five SNRs). The SRT and slope for each digit recording was read from the fitted parameters, enabling the most suitable digit for the test in each condition to be identified (the steepest slope), and level of each digit to be adjusted (based on the difference between the individual digit recording's SRT and the mean SRT of the selected digits). The results of this calculation are shown in tables 19 to 22 below. For all test conditions, the lowest slope of intelligibility chosen was 9.9%/dB and the highest was 82.1%/dB. The large difference is specifically observed for the digit 7 or /tujuh/ which contains unvoiced fricatives.

Table 20: Digit selection for headphone application using test specific noise

	Take 0					Take 1				
Position 1	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider
0	-15.45 dB	8.23 %/dB	0.0051	-2.06 dB	N	-14.45 dB	10.12 %/dB	0.0030	-1.06 dB	Y
1	-17.56 dB	18.48 %/dB	0.0011	-4.16 dB	Y	-19.03 dB	7.91 %/dB	0.0010	-5.63 dB	N
2	-12.86 dB	13.12 %/dB	0.0023	0.54 dB	N	-13.42 dB	20.92 %/dB	0.0000	-0.03 dB	Y
3	-13.58 dB	16.08 %/dB	0.0035	-0.18 dB	N	-13.99 dB	16.77 %/dB	0.0004	-0.59 dB	Y
4	-13.69 dB	28.70 %/dB	0.0001	-0.29 dB	Y	-13.39 dB	14.20 %/dB	0.0016	0.01 dB	N
5	-11.57 dB	21.21 %/dB	0.0018	1.83 dB	Y	-12.49 dB	11.01 %/dB	0.0005	0.90 dB	N
6	-11.99 dB	30.73 %/dB	0.0003	1.41 dB	Y	-12.90 dB	15.32 %/dB	0.0006	0.50 dB	N
7	-15.20 dB	23.45 %/dB	0.0001	-1.80 dB	N	-15.81 dB	25.64 %/dB	0.0000	-2.42 dB	Y
8	-11.04 dB	14.01 %/dB	0.0012	2.35 dB	Y	-11.80 dB	10.30 %/dB	0.0019	1.60 dB	N
Position 2										
0	-14.87 dB	10.61 %/dB	0.0024	-1.47 dB	N	-15.50 dB	12.91 %/dB	0.0042	-2.10 dB	Y
1	-16.87 dB	9.90 %/dB	0.0019	-3.48 dB	Y	-16.46 dB	9.07 %/dB	0.0021	-3.06 dB	N
2	-16.34 dB	16.89 %/dB	0.0007	-2.94 dB	Y	-16.39 dB	12.78 %/dB	0.0019	-3.00 dB	N
3	-12.67 dB	15.92 %/dB	0.0015	0.73 dB	Y	-13.28 dB	14.27 %/dB	0.0007	0.12 dB	N
4	-13.48 dB	17.82 %/dB	0.0023	-0.08 dB	Y	-13.27 dB	14.48 %/dB	0.0010	0.13 dB	N
5	-13.19 dB	14.65 %/dB	0.0009	0.21 dB	Y	-14.50 dB	10.76 %/dB	0.0004	-1.10 dB	N
6	-14.29 dB	23.26 %/dB	0.0009	-0.89 dB	Y	-13.45 dB	17.54 %/dB	0.0012	-0.06 dB	N
7	-16.51 dB	18.16 %/dB	0.0006	-3.11 dB	Y	-16.86 dB	15.88 %/dB	0.0025	-3.47 dB	N
8	-11.60 dB	92.91 %/dB	0.0019	1.79 dB	Y	-11.22 dB	17.93 %/dB	0.0027	2.18 dB	N
Position 3										
0	-14.05 dB	8.10 %/dB	0.0032	-0.65 dB	N	-15.81 dB	13.58 %/dB	0.0014	-2.42 dB	Y
1	-16.93 dB	9.04 %/dB	0.0030	-3.53 dB	N	-16.01 dB	11.30 %/dB	0.0008	-2.61 dB	Y
2	-13.44 dB	15.48 %/dB	0.0038	-0.04 dB	N	-13.86 dB	15.72 %/dB	0.0005	-0.47 dB	Y
3	-12.92 dB	14.89 %/dB	0.0012	0.47 dB	N	-14.28 dB	18.21 %/dB	0.0002	-0.89 dB	Y
4	-11.91 dB	24.04 %/dB	0.0016	1.49 dB	N	-11.74 dB	17.82 %/dB	0.0009	1.66 dB	Y
5	-13.93 dB	11.97 %/dB	0.0017	-0.53 dB	N	-12.46 dB	13.51 %/dB	0.0013	0.94 dB	Y
6	-10.98 dB	15.39 %/dB	0.0012	2.42 dB	N	-10.25 dB	21.21 %/dB	0.0055	3.14 dB	Y
7	-15.93 dB	22.70 %/dB	0.0003	-2.53 dB	N	-15.52 dB	23.74 %/dB	0.0001	-2.12 dB	Y
8	-12.36 dB	17.25 %/dB	0.0014	1.04 dB	N	-11.60 dB	20.76 %/dB	0.0008	1.80 dB	Y

Table 21: Digit selection for headphone application using spectrotemporal gap noise

	Take 0					Take 1				
Position 1	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider
0	-17.5 dB	7.75 %/dB	0.00083	**-2.6 dB**	Y	-16.9 dB	7.25 %/dB	0.00641	**-2.0 dB**	N
1	-18.1 dB	13.69 %/dB	0.00280	**-3.1 dB**	Y	-18.9 dB	11.29 %/dB	0.00327	**-3.9 dB**	N
2	-13.7 dB	11.05 %/dB	0.00177	**1.3 dB**	N	-13.4 dB	18.29 %/dB	0.00047	**1.5 dB**	Y
3	-15.0 dB	12.81 %/dB	0.00146	**-0.1 dB**	Y	-14.4 dB	8.27 %/dB	0.00148	**0.5 dB**	N
4	-13.5 dB	18.30 %/dB	0.00156	**1.4 dB**	Y	-14.0 dB	14.80 %/dB	0.00303	**0.9 dB**	N
5	-13.3 dB	10.67 %/dB	0.00028	**1.6 dB**	N	-13.7 dB	15.97 %/dB	0.00033	**1.3 dB**	Y
6	-13.0 dB	15.28 %/dB	0.00035	**1.9 dB**	Y	-13.7 dB	15.14 %/dB	0.00020	**1.3 dB**	N
7	-16.4 dB	22.83 %/dB	0.00132	**-1.4 dB**	N	-15.9 dB	26.47 %/dB	0.00171	**-1.0 dB**	Y
8	-13.2 dB	15.47 %/dB	0.00086	**1.8 dB**	Y	-14.0 dB	14.77 %/dB	0.00441	**0.9 dB**	N
Position 2										
0	-18.1 dB	10.42 %/dB	0.00044	**-3.1 dB**	N	-18.5 dB	24.66 %/dB	0.00191	**-3.6 dB**	Y
1	-17.4 dB	12.06 %/dB	0.00004	**-2.4 dB**	Y	-17.9 dB	9.65 %/dB	0.00052	**-3.0 dB**	N
2	-15.2 dB	11.19 %/dB	0.00386	**-0.2 dB**	N	-14.9 dB	13.88 %/dB	0.00058	**0.1 dB**	Y
3	-13.3 dB	15.56 %/dB	0.00051	**1.7 dB**	Y	-12.9 dB	15.55 %/dB	0.00344	**2.0 dB**	N
4	-12.9 dB	14.45 %/dB	0.00039	**2.1 dB**	Y	-13.9 dB	11.81 %/dB	0.00659	**1.0 dB**	N
5	-15.8 dB	17.72 %/dB	0.00017	**-0.8 dB**	Y	-16.0 dB	15.41 %/dB	0.00053	**-1.0 dB**	N
6	-14.3 dB	16.78 %/dB	0.00020	**0.7 dB**	N	-14.6 dB	22.20 %/dB	0.00035	**0.3 dB**	Y
7	-16.2 dB	15.83 %/dB	0.00088	**-1.2 dB**	N	-16.3 dB	26.25 %/dB	0.00025	**-1.4 dB**	Y
8	-12.6 dB	18.25 %/dB	0.00034	**2.3 dB**	Y	-12.9 dB	10.59 %/dB	0.00266	**2.0 dB**	N
Position 3										
0	-19.0 dB	16.94 %/dB	0.00265	**-4.1 dB**	Y	-17.9 dB	16.48 %/dB	0.00038	**-3.0 dB**	N
1	-18.5 dB	6.13 %/dB	0.00072	**-3.5 dB**	N	-17.8 dB	11.53 %/dB	0.00033	**-2.9 dB**	Y
2	-13.0 dB	15.81 %/dB	0.00042	**2.0 dB**	N	-13.1 dB	25.12 %/dB	0.00297	**1.8 dB**	Y
3	-13.5 dB	25.48 %/dB	0.00296	**1.4 dB**	Y	-14.1 dB	12.81 %/dB	0.00026	**0.8 dB**	N
4	-12.1 dB	12.23 %/dB	0.00036	**2.9 dB**	N	-12.2 dB	14.32 %/dB	0.00133	**2.8 dB**	Y
5	-15.9 dB	11.47 %/dB	0.00069	**-0.9 dB**	N	-15.1 dB	13.06 %/dB	0.00180	**-0.2 dB**	Y
6	-12.6 dB	12.55 %/dB	0.00145	**2.4 dB**	N	-11.3 dB	14.86 %/dB	0.00044	**3.6 dB**	Y
7	-16.6 dB	18.00 %/dB	0.00051	**-1.6 dB**	Y	-15.5 dB	97.67 %/dB	0.00090	**-0.5 dB**	N
8	-13.6 dB	15.80 %/dB	0.00245	**1.4 dB**	Y	-13.1 dB	14.80 %/dB	0.00107	**1.9 dB**	N

Table 22: Digit selection for telephone application using test specific noise

	Take 0					Take 1				
Position 1	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider
0	-15.2 dB	8.77 %/dB	0.0005	**-1.8 dB**	N	-15.9 dB	10.02 %/dB	0.00252	**-2.3 dB**	Y
1	-16.7 dB	13.41 %/dB	0.0008	**-3.3 dB**	Y	-17.4 dB	9.81 %/dB	0.00154	**-3.9 dB**	N
2	-12.1 dB	23.73 %/dB	0.0033	**1.3 dB**	Y	-12.6 dB	18.72 %/dB	0.00169	**1.0 dB**	N
3	-13.1 dB	13.90 %/dB	0.0003	**0.3 dB**	N	-13.0 dB	15.33 %/dB	0.00044	**0.6 dB**	Y
4	-12.5 dB	16.40 %/dB	0.0008	**0.9 dB**	Y	-14.2 dB	13.41 %/dB	0.00084	**-0.6 dB**	N
5	-12.2 dB	13.18 %/dB	0.0020	**1.2 dB**	N	-12.5 dB	13.45 %/dB	0.00095	**1.0 dB**	Y
6	-12.8 dB	15.11 %/dB	0.0018	**0.6 dB**	N	-12.8 dB	21.80 %/dB	0.00085	**0.8 dB**	Y
7	-15.4 dB	78.77 %/dB	0.0019	**-2.0 dB**	N	-15.2 dB	82.09 %/dB	0.00077	**-1.6 dB**	Y
8	-10.8 dB	10.10 %/dB	0.0023	**2.6 dB**	Y	-11.0 dB	8.51 %/dB	0.00308	**2.5 dB**	N
Position 2										
0	-15.8 dB	10.40 %/dB	0.0006	**-2.4 dB**	N	-16.0 dB	14.15 %/dB	0.00162	**-2.5 dB**	Y
1	-16.8 dB	15.57 %/dB	0.0012	**-3.4 dB**	Y	-16.6 dB	11.71 %/dB	0.00094	**-3.0 dB**	N
2	-16.5 dB	13.59 %/dB	0.0006	**-3.0 dB**	N	-16.4 dB	15.17 %/dB	0.00077	**-2.8 dB**	Y
3	-13.3 dB	14.41 %/dB	0.0040	**0.1 dB**	Y	-13.7 dB	16.36 %/dB	0.00705	**-0.1 dB**	N
4	-12.4 dB	14.82 %/dB	0.0017	**1.0 dB**	N	-12.9 dB	22.40 %/dB	0.00288	**0.7 dB**	Y
5	-13.8 dB	22.69 %/dB	0.0045	**-0.3 dB**	Y	-13.6 dB	12.59 %/dB	0.00059	**0.0 dB**	N
6	-14.7 dB	20.85 %/dB	0.0003	**-1.3 dB**	Y	-14.4 dB	19.96 %/dB	0.00118	**-0.8 dB**	N
7	-16.0 dB	35.66 %/dB	0.0017	**-2.6 dB**	Y	-16.7 dB	13.55 %/dB	0.00045	**-3.1 dB**	N
8	-11.8 dB	14.86 %/dB	0.0004	**1.7 dB**	Y	-11.7 dB	22.88 %/dB	0.00121	**1.9 dB**	N
Position 3										
0	-16.1 dB	13.67 %/dB	0.0009	**-2.7 dB**	Y	-16.3 dB	11.07 %/dB	0.00557	**-2.7 dB**	N
1	-16.4 dB	11.95 %/dB	0.0013	**-3.0 dB**	Y	-15.9 dB	9.81 %/dB	0.00090	**-2.3 dB**	N
2	-12.8 dB	21.32 %/dB	0.0003	**0.7 dB**	Y	-12.2 dB	15.50 %/dB	0.00145	**1.3 dB**	N
3	-14.3 dB	16.10 %/dB	0.0032	**-0.9 dB**	Y	-14.4 dB	12.31 %/dB	0.00028	**-0.8 dB**	N
4	-11.7 dB	25.55 %/dB	0.0030	**1.8 dB**	N	-12.0 dB	25.59 %/dB	0.00115	**1.6 dB**	Y
5	-12.5 dB	8.78 %/dB	0.0012	**1.0 dB**	N	-14.8 dB	14.15 %/dB	0.00193	**-1.3 dB**	Y
6	-11.3 dB	24.46 %/dB	0.0010	**2.2 dB**	Y	-10.8 dB	16.66 %/dB	0.00221	**2.8 dB**	N
7	-16.2 dB	20.31 %/dB	0.0003	**-2.7 dB**	N	-15.4 dB	75.84 %/dB	0.00247	**-1.8 dB**	Y
8	-12.7 dB	12.25 %/dB	0.0021	**0.7 dB**	N	-12.0 dB	13.31 %/dB	0.00297	**1.6 dB**	Y

Table 23: Digit selection for telephone application using spectrotemporal gap noise

	Take 0					Take 1				
Position 1	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider	Lmid	Slope at midpoint:	mserror	Lmid adjustment (dB)	Consider
0	-16.7 dB	9.19 %/dB	0.00343	**-2.5 dB**	Y	-16.5 dB	8.82 %/dB	0.00166	-2.2 dB	N
1	-17.5 dB	12.91 %/dB	0.00312	**-3.3 dB**	Y	-18.8 dB	7.78 %/dB	0.00140	-4.5 dB	N
2	-12.7 dB	19.09 %/dB	0.00150	**1.5 dB**	Y	-12.3 dB	15.53 %/dB	0.00353	2.0 dB	N
3	-14.1 dB	21.22 %/dB	0.00015	**0.1 dB**	Y	-14.0 dB	10.95 %/dB	0.00381	0.2 dB	N
4	-13.7 dB	13.57 %/dB	0.00155	**0.5 dB**	N	-12.6 dB	19.01 %/dB	0.00255	1.7 dB	Y
5	-14.6 dB	13.73 %/dB	0.00106	**-0.4 dB**	Y	-14.8 dB	13.53 %/dB	0.00209	-0.5 dB	N
6	-13.4 dB	24.09 %/dB	0.00034	**0.8 dB**	Y	-13.5 dB	15.79 %/dB	0.00013	0.7 dB	N
7	-15.8 dB	21.31 %/dB	0.00044	**-1.6 dB**	N	-16.8 dB	22.82 %/dB	0.00165	-2.5 dB	Y
8	-13.5 dB	9.61 %/dB	0.00153	**0.7 dB**	N	-12.7 dB	11.09 %/dB	0.00023	1.6 dB	Y
Position 2										
0	-17.6 dB	11.52 %/dB	0.00173	**-3.4 dB**	N	-19.0 dB	13.28 %/dB	0.00111	-4.7 dB	Y
1	-16.3 dB	13.43 %/dB	0.00093	**-2.1 dB**	Y	-17.2 dB	9.77 %/dB	0.00381	-2.9 dB	N
2	-15.5 dB	10.48 %/dB	0.01193	**-1.3 dB**	N	-14.3 dB	15.26 %/dB	0.00687	-0.1 dB	Y
3	-13.5 dB	21.12 %/dB	0.00608	**0.6 dB**	Y	-13.4 dB	18.21 %/dB	0.00018	0.9 dB	N
4	-11.8 dB	14.27 %/dB	0.00127	**2.4 dB**	N	-12.4 dB	14.52 %/dB	0.00660	1.9 dB	Y
5	-16.1 dB	20.53 %/dB	0.00061	**-1.9 dB**	Y	-16.5 dB	17.48 %/dB	0.00126	-2.2 dB	N
6	-14.8 dB	17.97 %/dB	0.00034	**-0.6 dB**	N	-14.5 dB	18.46 %/dB	0.00012	-0.2 dB	Y
7	-16.5 dB	27.16 %/dB	0.00266	**-2.3 dB**	Y	-16.8 dB	16.42 %/dB	0.00045	-2.5 dB	N
8	-13.0 dB	12.22 %/dB	0.00048	**1.2 dB**	N	-12.5 dB	22.47 %/dB	0.00041	1.7 dB	Y
Position 3										
0	-19.3 dB	10.21 %/dB	0.00012	**-5.1 dB**	Y	-19.2 dB	7.82 %/dB	0.00397	-4.9 dB	N
1	-15.9 dB	9.58 %/dB	0.00404	**-1.7 dB**	N	-16.8 dB	12.44 %/dB	0.00138	-2.5 dB	Y
2	-12.0 dB	24.55 %/dB	0.00157	**2.2 dB**	Y	-12.1 dB	20.48 %/dB	0.00102	2.2 dB	N
3	-14.5 dB	23.44 %/dB	0.00265	**-0.3 dB**	Y	-14.0 dB	12.62 %/dB	0.00224	0.3 dB	N
4	-11.7 dB	16.46 %/dB	0.00123	**2.5 dB**	Y	-12.0 dB	14.85 %/dB	0.00447	2.3 dB	N
5	-15.5 dB	15.20 %/dB	0.00477	**-1.3 dB**	Y	-15.1 dB	10.67 %/dB	0.00020	-0.8 dB	N
6	-12.0 dB	15.50 %/dB	0.00097	**2.2 dB**	N	-11.2 dB	18.01 %/dB	0.00098	3.1 dB	Y
7	-16.8 dB	15.63 %/dB	0.00012	**-2.6 dB**	N	-16.9 dB	18.80 %/dB	0.00013	-2.6 dB	Y
8	-12.7 dB	15.21 %/dB	0.00078	**1.5 dB**	N	-13.8 dB	20.06 %/dB	0.00088	0.4 dB	Y

Based on the results above, level adjustments were made to all digits to match mean SRTn for the selected digits in each tested condition. A summary of the adjustments are shown in the table below.

Table 24: Level adjustment figures for the MDTT in four testing conditions.

Test conditions	Average SRTn (dB SNR)	Pre-normalisation slope (%/dB)	Predicted slope (%/dB)	Average Level Adjustment (dB)	Minimum level of adjustment (dB)	Maximum level of adjustment (dB)
Headphone in TSN	-13.4	1.96	17.2	-0.5	-4	3.1
Headphone in STG	-14.4	1.43	16.5	-0.5	-4	3
Telephone in TSN	-13.5	1.54	21.4	-0.5	-3.3	2.7
Telephone in STG	-14.2	1.43	17.1	-0.5	-4	2.9

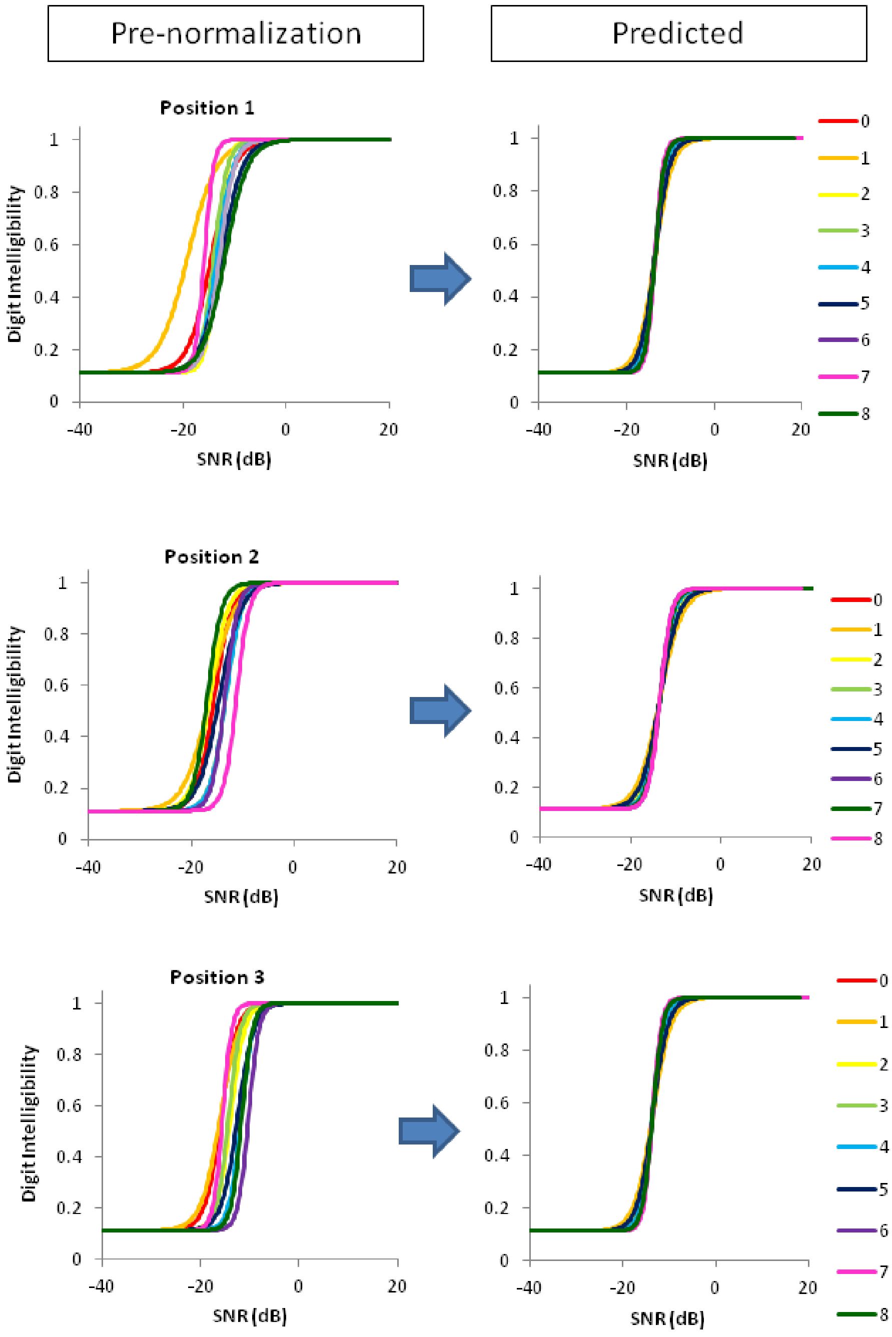


Figure 39: Psychometric function of digits using headphone in test specific noise before (left) and the predicted functions (right) after normalisation.

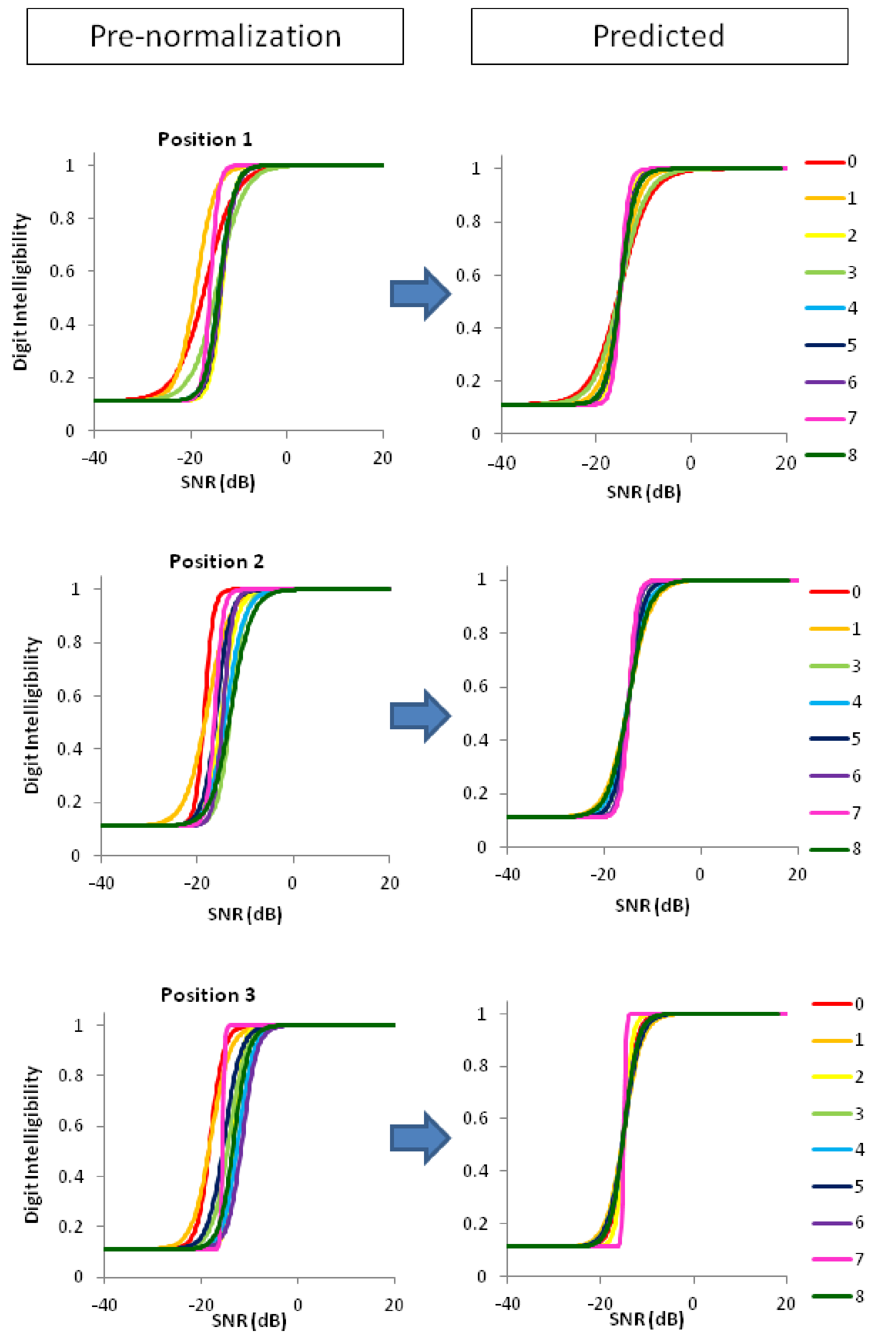


Figure 40: Psychometric function of digits using headphone and spectral temporal gap noise before (left) and the predicted functions (right) after normalisation.

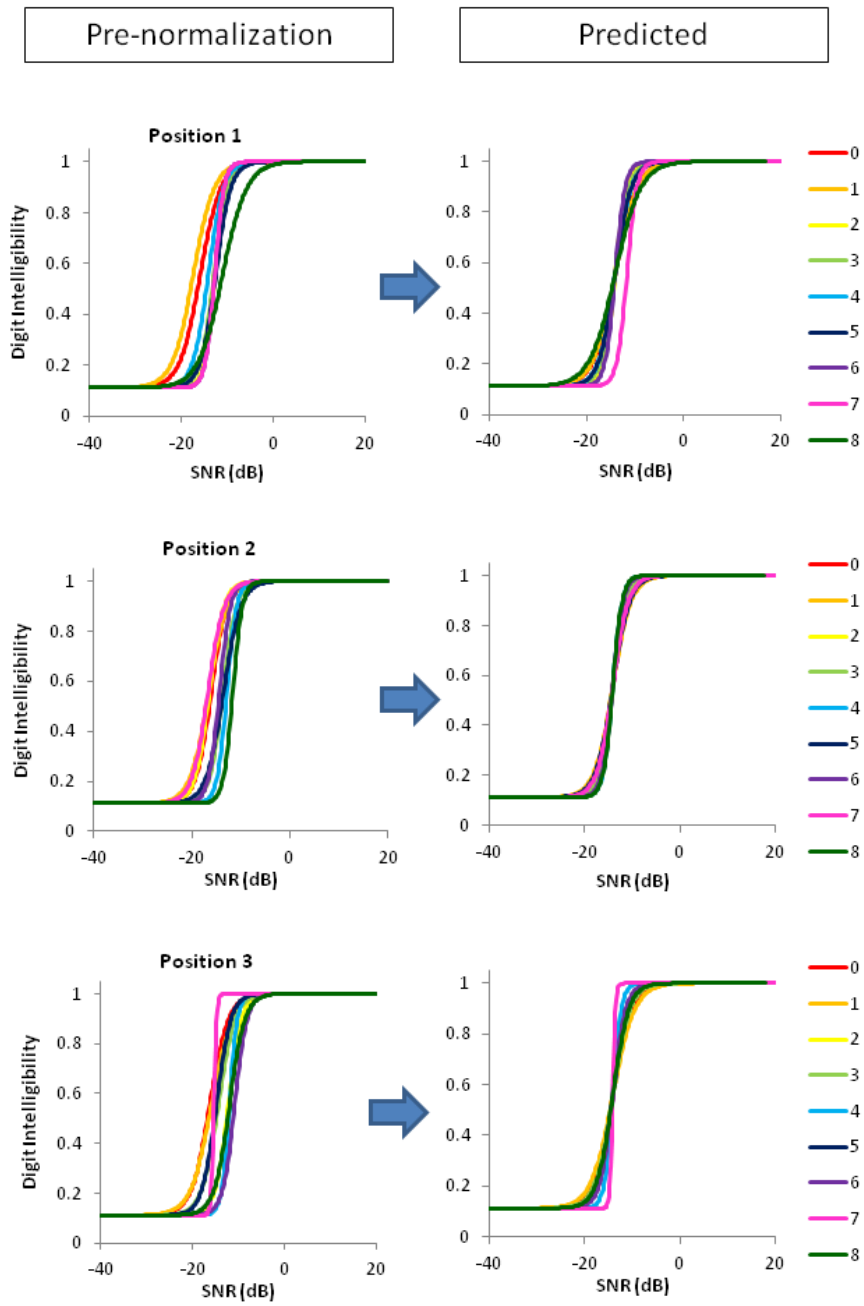


Figure 41: Psychometric function of digits using telephone and test specific noise before (left) and the predicted functions (right) after normalisation.

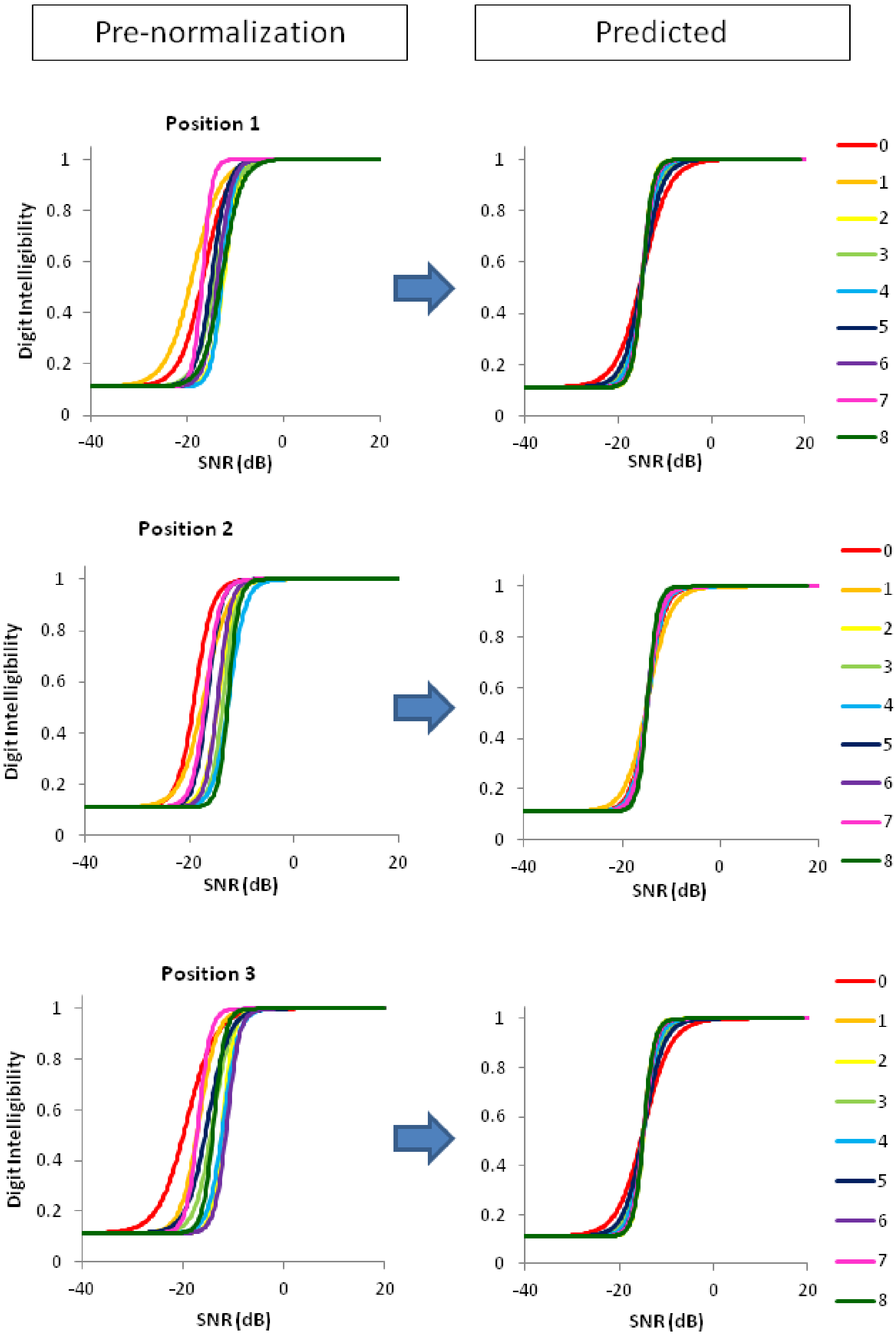


Figure 42: Psychometric function of digits using telephone and spectrotemporal gap noise before (left) and the predicted functions (right) after normalisation.

4.2.2.1 Constructing equivalent lists for the MDTT

The second phase of normalisation was to choose sets of triplets to form lists that could be evaluated and used in the final test. As none of the digits were rejected from the original recordings, all of the lists contained all digits in various unique triplet combinations for all four testing conditions. Eight equally intelligible test lists containing 27 digit triplets for each of the four test conditions were generated using custom-written software. As each generated list contained each digit exactly three times in each test position, the predicted mean SRTn for any list produced in this way would be identical. To ensure consistency in the slopes of the triplets in each generated list, the software used an iterative process that produced 2000 lists and selected the one with the smallest range of slopes. This was repeated until the required number of lists was generated such that no triplet appeared more than once across all lists.This ensured the distribution of mean triplet SRTn and slope were similar between lists. A summary of the test lists average SRTn and slope scores for all four conditions are shown in the table below.

Table 25: Summary of predicted test lists mean SRTn and slope scores for all four test conditions.

Test condition	Predicted mean SRTn (dB SNR)	Predicted mean slope (%/dB)	Predicted σ_{Slope} (%/dB)
Headphone in TSN	-13.39	18.4	2.7
Headphone in STG	-14.36	17.2	2.6
Telephone in TSN	-13.51	21.8	8.4
Telephone in STG	-14.21	17.8	2.6

The figures below demonstrate the distribution of triplet SRTn and slope scores for all four test conditions.

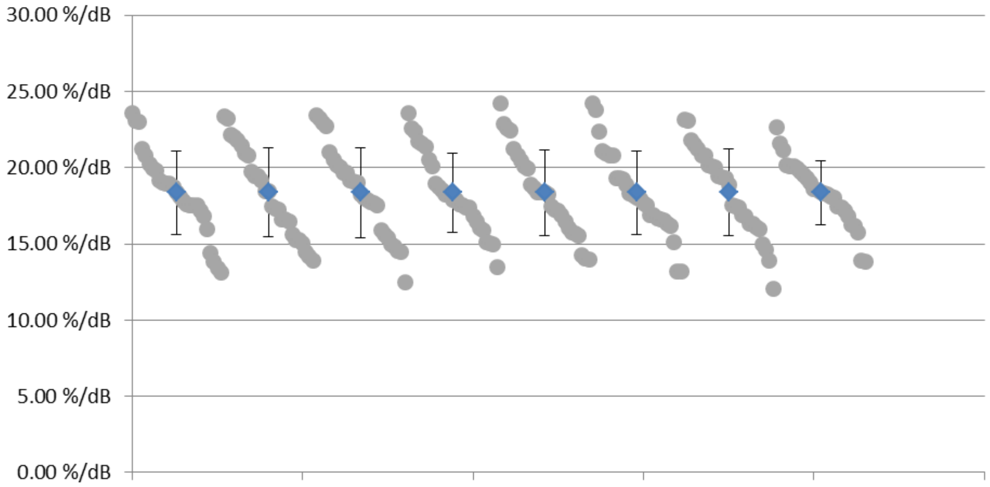


Figure 43: Distribution of triplet slopes for the MDTT lists for headphone application in steady state speech shaped noise (Lists 1-8).

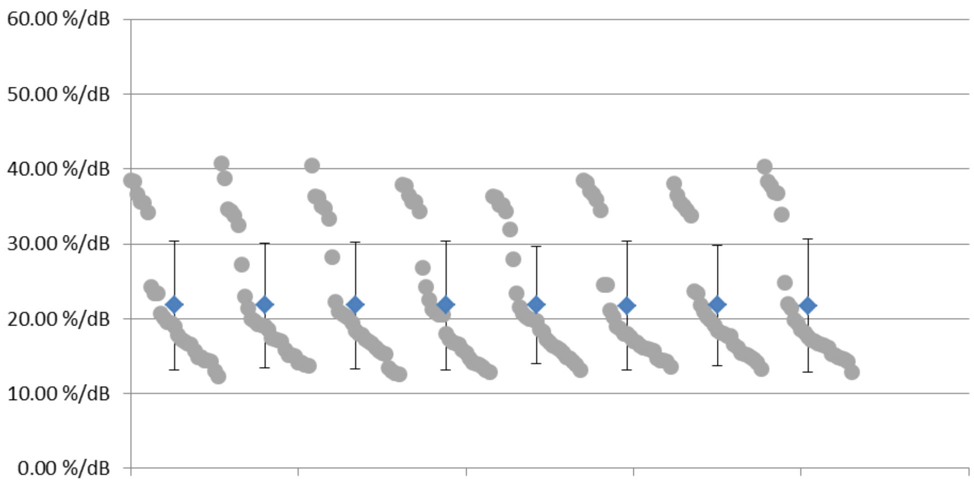


Figure 44: Distribution of triplet slopes for the MDTT lists for telephone application in steady state speech shaped noise (Lists 1-8).

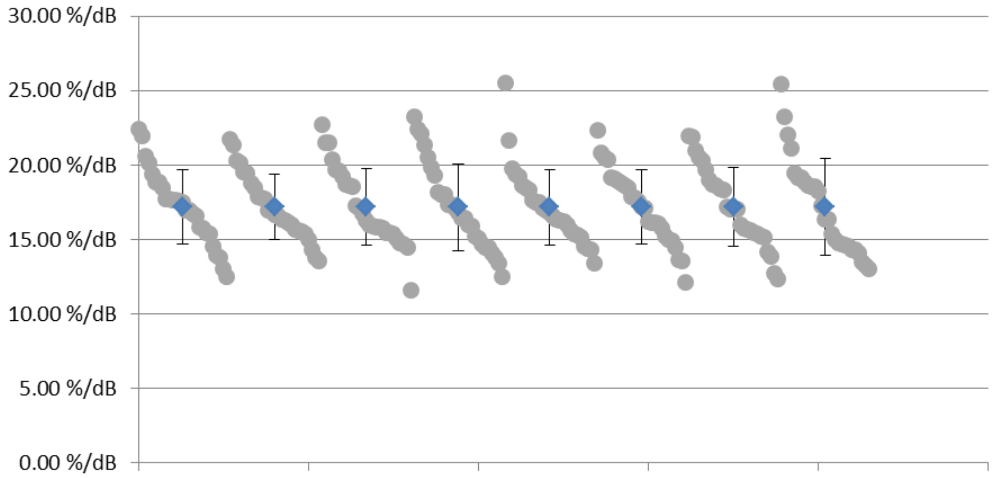


Figure 45: Distribution of triplet slopes for the MDTT lists for headphone application in spectrotemporal gap noise (Lists 1-8).

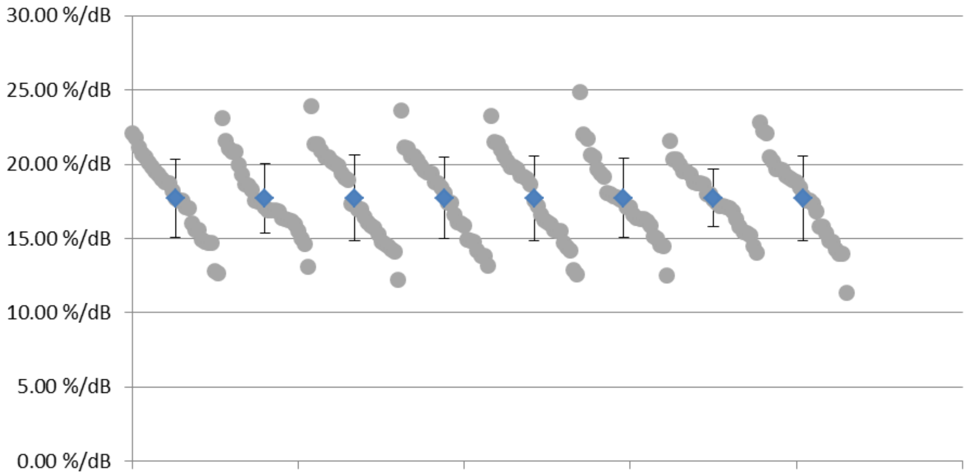


Figure 46: Distribution of triplet slopes for the MalayDTT lists for telephone application in spectrotemporal gap noise (Lists 1-8).

4.2.3 Discussion

4.2.3.1 Digit selection after normalisation

The recordings of Takes 1 & 2 allowed flexibility in choosing the appropriate digits as some recordings had either very low or very high slopes of intelligibility. As a general rule, digits that showed higher slope of intelligibility were chosen for the tests however some exceptions were made if digits showed artefactually high slope steepness. These digits were re-examined to investigate if the logistical function were fitted appropriately. If the function had shown a reasonable fit between the data and curve, the digit slope score was then considered as acceptable. For example in the figure below, the digit 7 from Take 1 for the MDTT using headphone in STG was selected instead of Take 2 as the logistical function fitted the data more accurately. The curve for digit 7 in Take 2 between -15 and -11.5 dB SNRs did not fit data sets making the curve steeper than it could reasonably be.

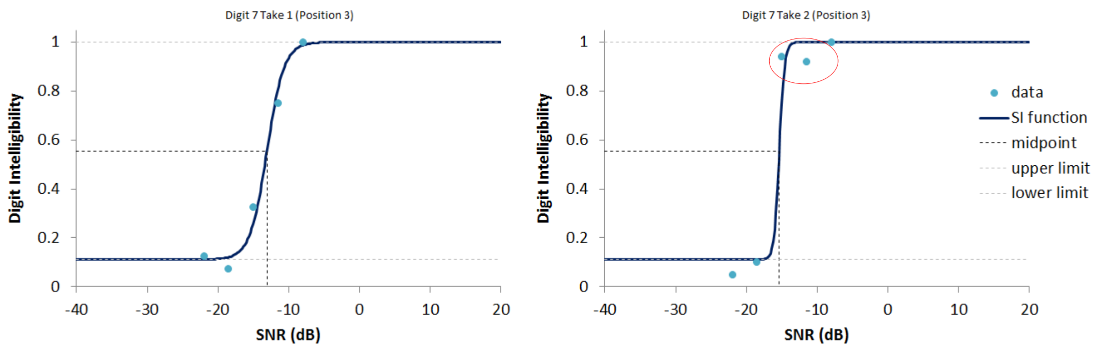


Figure 47: Comparison between slope of intelligibility of digit 7 of Takes 1 (18%/dB) & 2 (98%/dB) at the back or third position of the triplet using headphone in STG

The example below shows two highly steep curves for digit 7 for both takes of recording in the "telephone in TSN" condition. Both curves seem to fit reasonably well between data sets hence the digit from Take 2 was selected as it showed a steeper slope compared to Take 1.

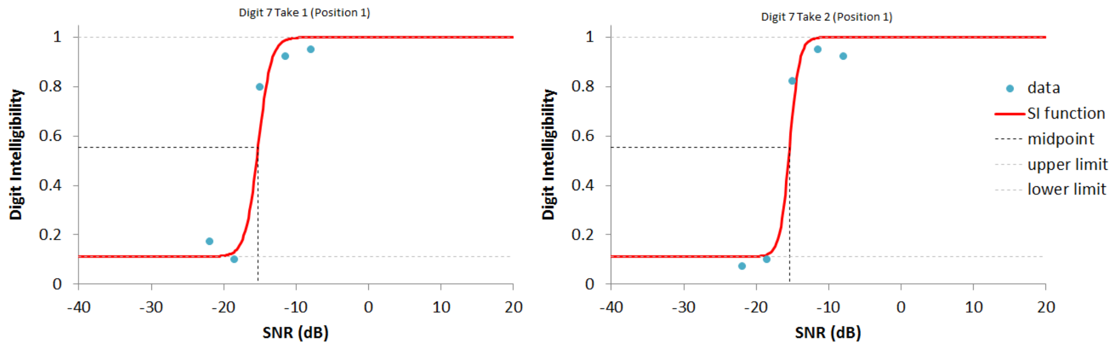


Figure 48: Comparison between slope of intelligibility of digit 7 of Takes 1 (79%/dB) & 2 (82%/dB) at the front or first position of the triplet using telephone in TSN.

Overall, no digits were excluded from the original digit selection as we took advantage of the 2 takes of recordings. The intelligibility slopes were examined to ensure the curve fitted the data reasonably. Digit 7 ("tujuh") consistently showed very steep slopes throughout all four conditions of testing. This is possible due to the phonetic composition of the digit which contains 3 unvoiced fricatives /t/, /ʧ / and /h/ sound combined with the vowel /u/ which is a low back vowel (Najibah Mahirah Awang, 2015). The combination of these phonemes may have resulted in higher dependency on the audibility of the vowel /u/ with minimum contribution from the voiceless consonants making it more difficult for listeners to hear.

4.2.3.2 Improvement of slope

The normalisation process revealed large improvements of the predicted slope of intelligibility after normalisation. An average 15.75%/dB of slope improvement was seen across all four test conditions. Using telephones showed higher improvements (15%/dB) compared to headphones (18.9%/dB). There were no attempts made to include more digits that contained higher slopes to improve list intelligibility as the lists were designed to be uniform in their distribution of digits. Increasing the frequency of digits with high slopes may lead to higher redundancy in the triplets that could cause listeners to memorize or expect the digits more (Houben & Dreschler, 2015). No inferential analysis was attempted to compare between the effects of using different transducers and background noise at this stage of the study as the normalisation process will be evaluated in a different group of normal hearing listeners which will bedescribed in Chapter 7.

4.2.3.3 MDTT list construction

The iteratively generated lists showed overall consistency in terms of the predicted slope for all test conditions and the inclusion of unique triplet between lists retained equivalent SRTn values between lists when compared to the overall slope scores of individual digits after normalisation. Digit lists showed overall difference of 0.75%/dB between lists-specific slopes and digit-specific slopes. All digits were equally represented in all lists and all four test conditions making the test phonemically balanced as well.

CHAPTER 5

NORMALISATION AND REFINEMENT OF THE MALAY AUDITORY-VISUAL MATRIX SENTENCE TEST.

5.1 Introduction

As discussed in the previous chapter, the process of normalisation using psychometric theory is key to improving test sensitivity speech tests. This process allowed the established MSTs to have considerably higher slopes of intelligibility compared to the HINT test (Soli & Wong, 2008) even though the test material is highly redundant in its structure. The use of homogenous interchangeable words that has been optimized in the MST made the test highly unpredictable, hence producing the steep slopes. Test with very steep slopes have high degree of sensitivity which allows for better precision and is robust against errors – even when using small number of trials. The main objective of this study was to normalise the words used in the MMST-AV in normal hearing subjects. To recruit an adequate number of normal hearing Malay-speaking subjects, this study was conducted in Malaysia. Normalisation was attempted in auditory-alone mode using headphones in two types of background noise: test specific noise or test-specific noise (TSN) and 6-talker babble noise (BN). An additional study was done to evaluate the quality of the video recording edits for the MMST-AV.

As with other MSTs, the words in the MMST were recorded in continuous sentences and then edited in word pairs to preserve the natural speech transitions. This is to allow more natural sounding sentences when different word combinations were used. This technique was first introduced in the Danish MST (Wagener et al., 2003) and is currently the reference technique of editing for other versions of the MST. As this study attempted to develop an auditory-visual version of the MST, recording and editing techniques had to be revised to ensure both media would show consistency when it is played to the listener. As discussed in Chapter 3, a few words selected for this test were edited between syllables to maintain natural speech transitions as well continuous video recording. As this editing process does not allow the usual word scoring method, Trounson (2012) proposed an alternative scoring method to resolve this issue (see Chapter 3). As some of the audio files were edited between syllables, a word playback of stimuli could be generated from an entire or partial audio files therefore two possible methods of normalisation were considered. The first possible method of normalisation can be achieved by level adjusting actual sound file fragments and alternatively, based on the scoring method proposed by Trounson (2012), specific word level adjustments can be made to produce the same result. To identify the best method for normalisation for this study, normal hearing listeners were evaluated in both fragment and word normalisation methods – these two methods are described in depth in Section 5.2.2.3 below.

Finally, the quality of video transitions between video fragments was evaluated. The aim of this part of the study was to identify the level of pixel difference at which viewers found the transitions to be as natural as the original continuous sentences, or at least not distracting. Video fragments that contained significantly higher levels of movements in frame or “judder” were to be excluded from the final version of the test.

5.2 Methods

5.2.1 Study 1: Normalisation of speech stimuli

5.2.1.1 Sample size calculations

As with the Malay DTT (described previously in Section 4.2.1.1), in order to study repeated measures of the MST for normalisation purposes, the sample size of the each group was calculated using Power & Sample size calculator software (Dupont & Plummer, 1990). The target power (the ability to reject the null hypothesis in favour of a specific true alternative) was set to a probability of 0.9 - the Type I error probability associated with this test of this null hypothesis was 0.05. In a previous study (Wagener et al., 2003) the response within the normal hearing group was normally distributed with standard deviation of less than 0.2 dB, which indicates that at least 9 normal hearing (HTL 0.5, 1, 2, 4 kHz < 20 dB HL) subjects should be recruited to achieve the predicted power of study.

5.2.1.2 Recruitment of participants

A total of 36 Malay native speakers aged 19 to 41 were recruited for this study (average age ± S.D. = 27.5 ± 6.3 years). Nineteen participants completed the normalisation study in TSN and the rest were tasked to complete the normalisation process for the MMST in BN. All participants were recruited from the faculty and students of the Allied Health Faculty, International Islamic University Malaysia in Kuantan, Malaysia. All participants had normal hearing (HTL 0.5, 1, 2, 4 kHz < 20 dB HL) and no history of ear or balance problems. All participants were paid RM50 (~NZD 17) for their time and effort.

5.2.1.3 Test procedures

During the speech perception measurements, participants were seated in a double walled audiometric cabin and were asked to choose words that were heard using a closed-set response method using the computer interface shown in Figure 49 below. For the purpose of normalizing the auditory stimuli, the test was presented without any visual cues so as to eliminate the contribution of lip-reading. The stimuli were presented via an external sound card (Sound Blaster X-Fi Surround 5.1 Pro, Creative Labs, Singapore) and Sennheiser HD 280 Pro headphones (Sennheiser electronic GmbH & Co., Germany). All participants were briefed about the test and given 20 practice sentences before the actual test commenced.

The practice runs were to familiarize the participants to the test and avoid any training effect (Wagener et al., 2003).

Presentation of test was done monaurally and kept at 65 dB A and target SNR levels were matched by adjusting the intensity of sentence presentation. For the test in TSN, participants were then asked to respond to 500 sentences (5 SNRs (-15.2, -11.7, -8.2, -4.7 & -1.2 dB SNR) *100 sentences containing all 400 possible pairs) whereas the listeners tested in BN were asked to respond to 700 sentences (7 SNRs (-17.9, -14.4, -10.9, -7.4, -3.9, -0.4 & 3.1 dB SNR) *100 sentences containing all 400 possible pairs).. Participants that were enrolled for the test using BN were asked to complete the task at 2 additional SNRs because the responses obtained from the first few participants did not reach the targeted 80% score at the highest SNR level which could have affected the overall morphology of the intelligibility function. Participants took an average of 3 hours and 6 minutes to complete the test in TSN and 3 hours 44 minutes for the experiment in BN. Participants were allowed rests in between sentences by using a pause button in the software interface.

Figure 49: Malay matrix normalisation interface, shown after the participant had entered four of the five words of their response.

5.2.1.4 Determining SRT levels

To determine the word and fragment-specific intelligibility functions, speech perception was assessed at fixed signal-to-noise ratios using the following function.

Equation 3: Word specific intelligibility function as described in Kollmeier et al., (2015). With, SRT_{word}: words-specific SRT in dB and S_{50word}: Slope at the SRT_{word} in 1/dB.

Word and fragment specific scores were averaged across all speech samples and used as basis of level adjustments for the test in both noises. The corrections were be applied in the next study, which was to identify the most appropriate method of normalisation for this test.

5.2.2 Study 2: Evaluation of possible normalisation procedures.

The aim of this study was to evaluate and identify the most appropriate normalisation method for the MMST-AV using the scoring method as proposed by Trounson (2012) based on the normalisation data obtained from Study 1.

5.2.2.1 Recruitment of participants

Nine normal hearing Malay native speakers (average hearing threshold levels at 250, 500, 1000, 2000, 4000 & 8000 Hz < 20 dB HL) were recruited at this stage of the study using convenience sampling. Participants were aged 24 to 38 years (average 28.6 ± 9.9 years) and had no history of ear and balance problems. Participants were Malaysian postgraduate students at the University of Canterbury and Lincoln University. All participants were given NZD 20 shopping voucher for their time and effort.

5.2.2.2 Testing procedure

All participants were tested in a single-walled audiometric cabin at Department of Communication Disorders research facility, Level 8, Rutherford Building, University of Canterbury, New Zealand. After completing audiometric threshold assessment, participants were asked to listen to a list of sentences normalised using both fragment and word methods in auditory mode only. The background noise chosen for this test was the steady state speech-shaped noise or test specific noise (TSN) because it provided energetic masking and, more importantly, because it is known to produce least variability in listeners response (Hochmuth, Kollmeier, et al., 2015; Wagener & Brand, 2005). Participants were tested at two SNRs (-10.2 and -6.3 dB SNR) using a list of 200 sentences (100 sentences * 2 SNRs) containing all possible word combinations. These two SNRs were chosen as they matched 40% and 80% target intelligibility levels that would enable fitting of the psychometric function and determination of the SRT and slope. These targets differ slightly from the 20% and 80% "pair of compromise" determined by Brand & Kollmeier (2002) to be optimal for estimating these two parameters – the lower of the two SNR levels was increased slightly to make the lengthy task less demoralising for the participants. Participants were asked to use a mouse to click their responses using the closed-set response method shown in Figure 49 above. During sentence presentation, the screen containing the word options was blanked to ensure participants concentrated on the listening task instead of browsing for words. Before participants were scored, they were given 20 practice sentences in order to remove any training effect during the actual test.

5.2.2.3 Fragment vs. word normalisation

The two types of normalisation technique are possible due to the editing technique and coding of the software. Instead of using pre-recorded and synthesized sentences, the software is able to form any sentence in real-time from a library of edited media files by concatenating chosen words to form a sentence. Border identification for word normalisation of were recorded based on a custom VI written by Assoc. Prof. Greg O'Beirne which enabled the experimenter to split all fragments into two halves at a point of their choosing (usually zero amplitude or onset of new syllable).The waveform of the fragment (for example, abang_dapat.wav) was displayed on the screen, and the experimenter moved a cursor to divide the fragment such that the components of "abang" were on the left of the cursor and those of "dapat" were on the right. Two buttons enabled the experimenter to play the audio portions separately to confirm the accuracy of the divisions, and a third saved the sample number that signified the border between the two words. This number was then used in the UCAST software to apply the required normalisation corrections in real time to the appropriate parts of each fragment.

The figure below demonstrates the normalisation technique using word normalisation and fragment normalisation.

Average word and fragment level adjustments values were obtained from Study 1 for the MMST auditory mode in TSN. The SRTn for both methods of normalisation were derived using Equation (3).

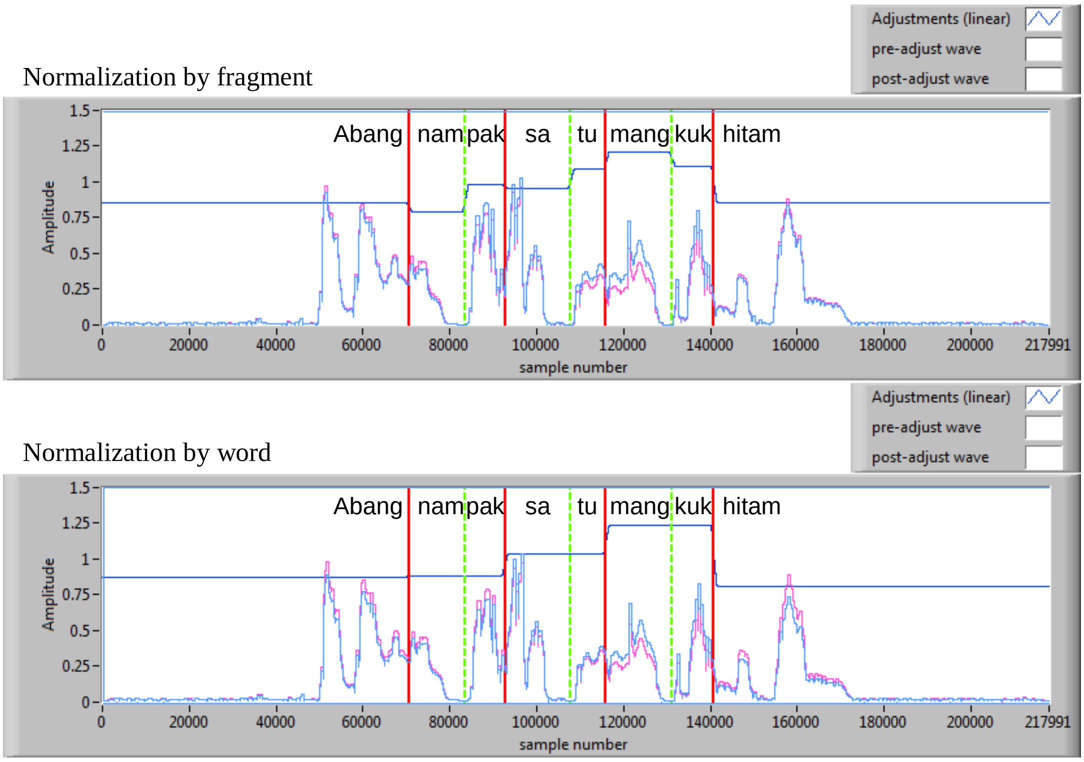


Figure 50: Demonstration of fragment and word normalisation for the sentence "Abang nampak satu mangkuk hitam" or *Brother sees one black bowl*. The figures show the pre- (pink line spectrum outline) and post-normalisation (light blue spectrum outline) for the sentence normalised based on fragment-specific score (top figure) and the same sentence normalised based on word-specific scores (bottom figure). In the word normalisation method, level adjustment could occur in between words as marked by the dotted green line whereas for the fragment normalisation, level adjustment occurs both between and within between recorded fragments marked by the green dotted and red lines. The blue horizontal stepped line demonstrates the amount of level adjustments (as a linear multiplier) for both types of normalisation.

5.2.3 Study 3: Refinement of the Malay auditory-visual matrix sentence test

The New Zealand English version of the auditory-visual matrix test was recorded in a controlled environment where various attempts were made to ensure the quality of the video is preserved (Trounson & O'Beirne, 2012). However, a review by 17 native New Zealand English speakers found significantly noticeable judders between many of the video fragments (McClelland, 2015). A reduction in the number of acceptable video transitions affects the total number of sentences that are available in the final version of the test. This also directly affects the word distribution between lists, hence limiting the number of possible test lists that can be used in the auditory-visual mode of the test. The aim of the present study was to determine the acceptable level of judder allowable in fragments that were to be used in the auditory-visual mode of the Malay version of the MST.

5.2.3.1 Method

5.2.3.2 Recruitment of participants

Ten adult native Malay speakers (6 males; 4 females) aged between 26 to 35 years old were recruited to rate the video quality of the MMST-AV. Participants assumed that they had normal hearing and vision were either normal or normal with optical correction. Participants were tested at the Department of Communication Disorders research facility at 19 Creyke Road, Ilam, Christchurch. All participants were postgraduate students at the University of Canterbury and were given NZD 20 shopping vouchers as compensation for their time and effort.

5.2.3.3 Testing procedure

All possible video transitions between fragments were examined by comparing the differences in pixel between the last and first frames of video. This produced a possible 3000 video transitions with varying levels of pixel difference. To be able to group the level of judders systematically, all possible judders were compared to the judders produced by the original 100 continuous recorded sentences, as shown in the figure below. Four categories of video judders called judder tiers were identified. Tier 1 judder was a conservative estimate based on one standard deviation of the average amount of judder produced by the original continuous sentences which is any video fragment judders less than 650,000 pixel difference. Tier 2 was labelled based on judders of video fragments between 650,000 and 975,000 pixel difference. Tiers 3 and 4 were based on judders of video fragments between 975,000 and 1,300,000, and 1,300,000 and 1,625,000 pixel difference respectively. Any judders over 1,625,000 pixel difference were rejected outright.

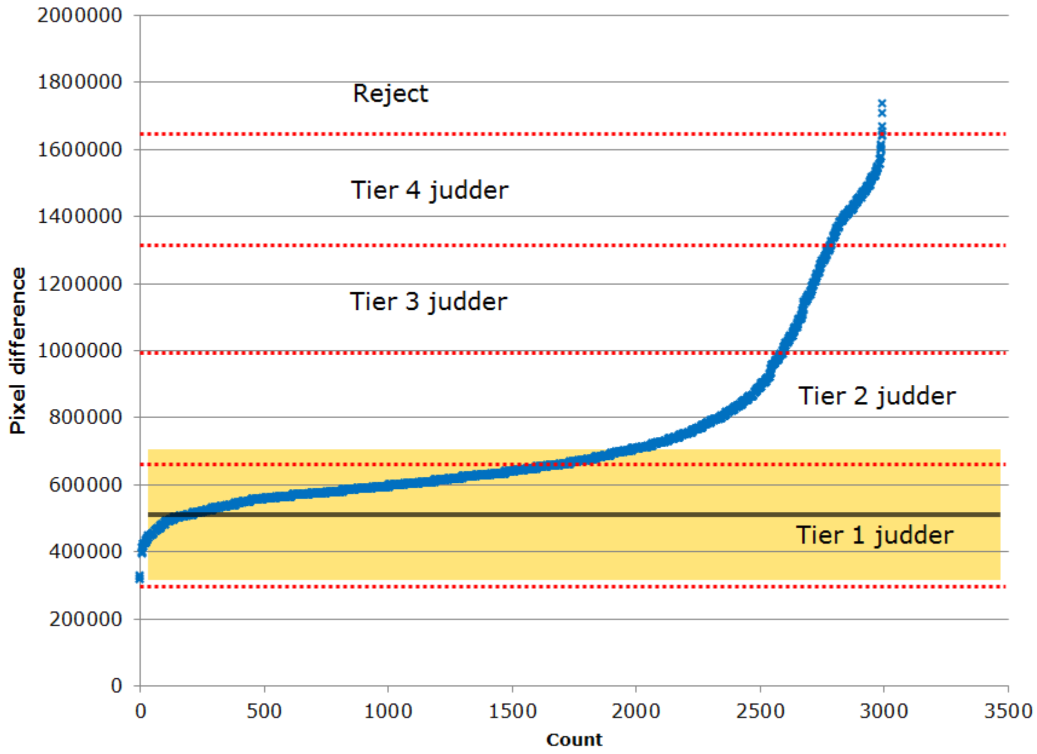


Figure 51: Distribution of the pixel difference between video frames of the 3000 possible video fragment transitions, ordered from smallest to largest. The black solid line and yellow shaded area indicates the mean and standard deviation of the amount of judder produced by the original 100 continuous unedited sentences (522,870 pixel difference ± 168,527).

Based on the possible judder categories, 600 sentences were generated that contained either 1 or 2 judders per sentence with varying judder tiers. The 600 sentences were then arranged in 10 lists containing 60 unique sentences each. The transitions of sentences with two judders were selected so as to had have pixel difference values within ± 2% of each other in the respective tier, preventing one more severe judder transition from influencing the rating score. A shorthand notation for describing the magnitude and position of the judders (described in Table 26 below) was developed for ease of analysis. The table also shows the distribution of the frequency of judders and judder tiers within a list.

Table 26: Distribution of judders and judder tiers per test list for the naturalness rating task.

Judder label	Description	Count
No judder	Sentences from original 100 continuous sentences	15
J1Tr01Ti1	One tier 1 judder at fragment position 1	2
J1Tr01Ti2	One tier 2 judder at fragment position 1	2
J1Tr01Ti3	One tier 3 judder at fragment position 1	2
J1Tr01Ti4	One tier 4 judder at fragment position 1	2
J1Tr02Ti1	One tier 1 judder at fragment position 2	2
J1Tr02Ti2	One tier 2 judder at fragment position 2	2
J1Tr02Ti3	One tier 3 judder at fragment position 2	2
J1Tr02Ti4	One tier 4 judder at fragment position 2	2
J1Tr03Ti1	One tier 1 judder at fragment position 3	2
J1Tr03Ti2	One tier 2 judder at fragment position 3	2
J1Tr03Ti3	One tier 3 judder at fragment position 3	2
J1Tr03Ti4	One tier 4 judder at fragment position 3	2
J2Tr12Ti1	Tier 1 judders at fragment positions 1 & 2	2
J2Tr12Ti2	Tier 2 judders at fragment positions 1 & 2	2
J2Tr12Ti3	Tier 3 judders at fragment positions 1 & 3	2
J2Tr12Ti4	Tier 4 judders at fragment positions 2 & 4	2
J2Tr13Ti1	Tier 1 judders at fragment positions 1 & 3	2
J2Tr13Ti2	Tier 2 judders at fragment positions 1 & 3	2
J2Tr13Ti4	Tier 4 judders at fragment positions 1& 3	2
J2Tr23Ti1	Tier 1 judders at fragment positions 2 & 3	2
J2Tr23Ti2	Tier 2 judders at fragment positions 2 & 3	2
J2Tr23Ti3	Tier 3 judders at fragment positions 2 & 3	1
J2Tr23Ti4	Tier 4 judders at fragment positions 2 & 3	2
	Total sentences	60

A custom VI was written by Assoc. Prof. Greg O'Beirne for the purpose of rating judders for this study. Before the task commenced, an introduction and instruction of this study was included, and examples of "no noticeable judder" and "highly noticeable judder" videos were presented to participants (see Figure 52 below for an the instructions in Malay given to participants). Sentences within the lists were also shuffled to avoid any learning effect. Participants used a 10-point sliding scale from "no noticeable judder" at 0 to "highly noticeable judder" at 10. The videos were presented together with sound so participants would also evaluate the edit points both auditory and visually. The audio signals were not normalised for this study and were presented via Sennheiser HD 280 Pro headphone. Participants were then presented with two lists each in sequence (Lists 1 and 2 for participant 1 and lists 3 and 4 for participant 2 and so on). The two lists were used to examine consistency of rating between to test trials.

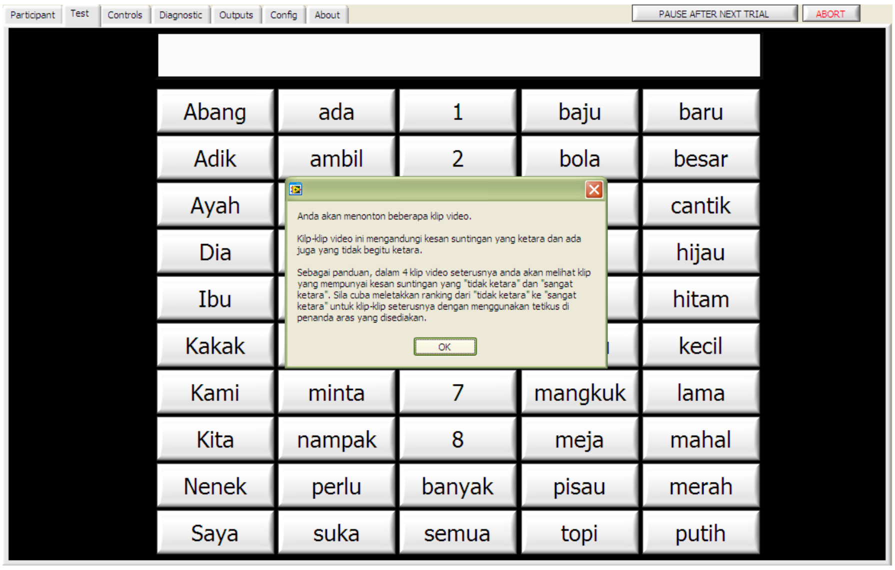


Figure 52: Instruction given prior to judder rating task. The English translation of the message reads: "You will be shown several video clips. Some video clips contains edits that will appear smooth while others contain noticeable judders. As a guide, the following 4 video clips will demonstrate videos with no noticeable judder and highly noticeable judder alternately. Please rank then from "no noticeable judder" to "highly noticeable judder" by clicking the on the measuring scale"

5.3 Results and analyses

5.3.1 Study 1: Normalisation of speech stimuli

The results described below were obtained for the purpose of investigating the average SRTn and the required level of adjustments so that they can be used to evaluate which of the two normalisations methods (word or fragment) were most appropriate for the MMST-AV. Further description of the chosen type of normalisation will be later discussed in this chapter together with findings of studies 2 and 3.

5.3.1.1 Normalisation in test specific noise (TSN)

5.3.1.1.1 Word normalisation

The average SRTn calculated from the data obtained in Study 1 using the word normalisation method was --9.03 dB with a standard deviation of 2.28 dB. The pre-normalisation slope of intelligibility word average was 13.4%/dB and the predicted post normalisation slope was 18.5%/dB. A limit of ± 3 dB level adjustment was set to ensure the audio signal sounded natural after normalisation, and so the average level adjustment for word normalisation was -0.04 dB ± 1.8. Using this ± 3 dB limit, 4 words were found to be capped at -3 dB and 5 words needed to be capped at 3 dB. The word "kecil" (or small) was found to be the easiest detectable word at -13.6 dB SNR whereas the most difficult word to perceive in this configuration of test was the word "bagi" (or give) at -2.9 dB SNR.

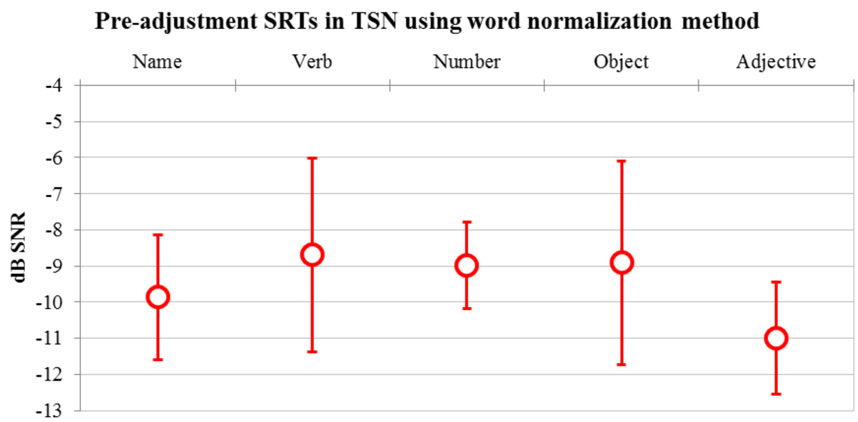


Figure 53: Average pre-normalisation SRTn scores of word categories calculated from the data obtained in Study 1 using the word normalisation method in TSN.

5.3.1.1.2 Fragment normalisation

The average SRTn calculated from the data obtained in Study 1 using the fragment normalisation method was -8.9 dB with a standard deviation of 2.2 dB. The slope of intelligibility was 11.2%/dB before normalisation and the predicted post-normalisation slope was 15.9%/dB. The same amount of level adjust limit was applied to this method of normalisation. This led to a total of 69 out of the 400 fragments that had to be capped at either 3 or -3 dB of level adjustment. The fragment “suka_banyak” had the lowest average SRTn of the fragments at -19.1 dB SNR, whereas the most detectable fragment was “bagi_dua” at -4.3 dB SNR.

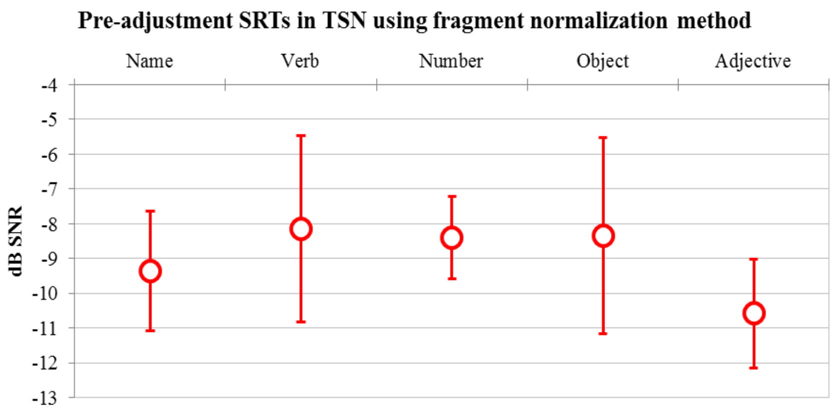


Figure 54: Average pre-normalisation SRTn scores of word categories calculated from the data obtained in Study 1 using the fragment normalisation method in TSN.

5.3.1.2 Normalisation in 6-talker babble noise (BN)

5.3.1.2.1 Word normalisation

The average SRTn in babble noise calculated from the data obtained in Study 1 using the word normalisation method was -4.65 dB with a standard deviation of ± 1.5 dB. The slope of intelligibility was 11.2%/dB before normalisation and the predicted post-normalisation slope was 14.1%/dB. We applied the same rule for level adjust limit which led to a total of 4 words that had to be capped at either 3 or -3 dB of level adjustment.

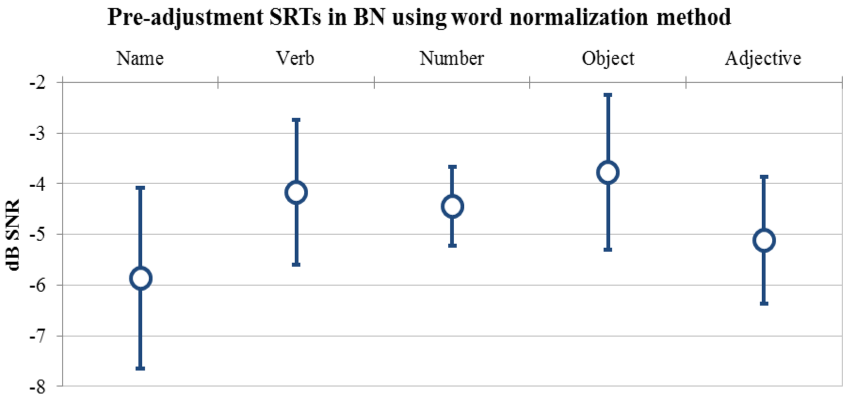


Figure 55: Average pre-normalisation SRTn scores of word categories calculated from the data obtained in Study 1 using the word normalisation method in BN.

5.3.1.2.2 Fragment normalisation

The calculations from the data obtained in Study 1 using fragment normalisation showed an average SRTn for all fragments of -4.63 dB SNR with standard deviation of ± 1.83 dB. The slope of intelligibility was predicted to be -14.8%/dB after normalisation, which is an increase from 13.6%/dB before normalisation. At total of 41 fragments that were capped at either 3 or -3 dB for level adjustment as their average fragment score was over the set adjustment a limit. The highest SRTn average was recorded for the fragment "buku_putih" at 0.38 dB SNR whereas the lowest SRTn was observed in the fragment "dia_mahu" at -12.24 dB SNR.

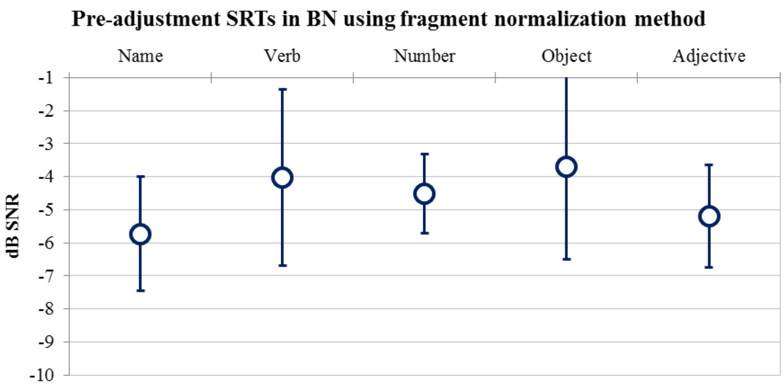


Figure 56: Average pre-normalisation SRTn scores of word categories calculated from the data obtained in Study 1 using the fragment normalisation method in BN.

5.3.2 Study 2: Evaluation of possible normalisation procedures.

Evaluation in nine normal hearing Malay speakers using both these normalisation methods revealed average scores for word and fragment normalisation in TSN at -10.2 dB SNR of 45% ± 0.11 and 44.7% ± 0.22 respectively. At -6.3 dB SNR, the score for word normalisation was 83.1% ±1.1 and for fragment normalisation it was 81% ±0.19. As shown in Figure 57 below, the range of the data was higher using fragment normalisation when compared to word normalisation. At -6.3 dB SNR, fragment normalisation resulted in a spread of almost 40% in score which included an outlier of item 6 – the word "bagi". The spread of data was even higher using fragment normalisation at -10.2 dB SNR where the spread of data was well over 80%.

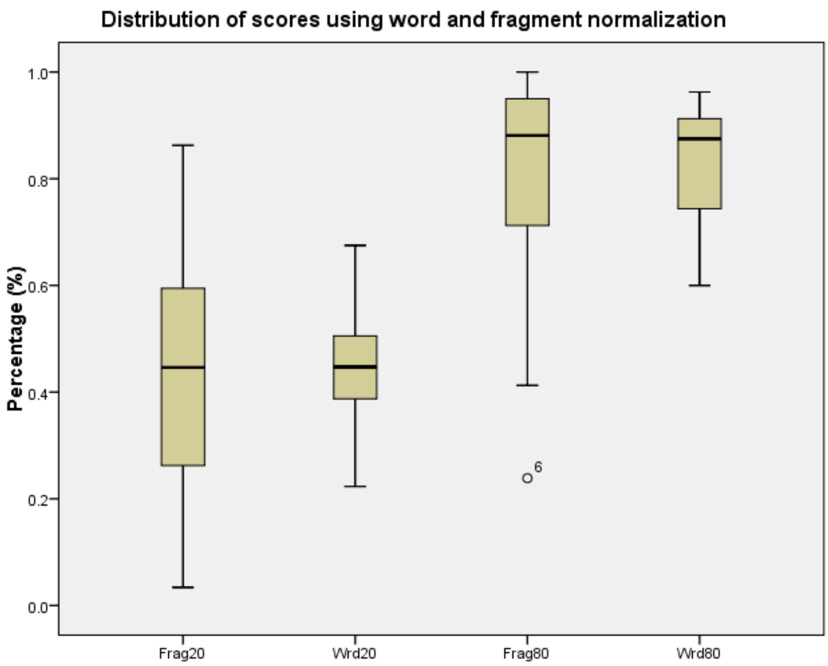


Figure 57: Boxplot of data distribution for word and fragment normalisation scores at -10.2 (Frag20 & Wrd20) and -6.3 dB (Frag80 & Wrd80) SNRs.

A descriptive analysis of the SRTn and slope measurements for both types of normalisation method is shown in Table 27 below.

Table 27: SRTn and slope of intelligibility for word normalisation (SRT_{WN} & $Slope_{WN}$) and fragment normalisation (SRT_{FN} & $Slope_{FN}$).

	Mean	Std. Deviation	Minimum	Maximum
SRT_{WN} (dB SNR)	-9.45	1.31	-11.58	-6.87
SRT_{FN} (dB SNR)	-9.48	2.17	-15.78	-4.71
$Slope_{WN}$ (%/dB)	11.55	3.13	5.32	17.95
$Slope_{FN}$ (%/dB)	13.95	5.66	5.69	41.28

There was no significant effect on SRTn ($t(43) = 0.247$, $p = 0.806$) observed for both methods of normalisation using a paired T-test. However a significant effect for slope of intelligibility ($t(43) = 2.646$, $p < 0.05$) was found between the two methods where $Slope_{FN}$ was higher than $Slope_{WN}$ by 1.5%/dB. $Slope_{FN}$ was closer to the predicted slope score using fragment normalisation (difference of 0.5%/dB) whereas $Slope_{WN}$ was different by 1.4%/dB when compared to the predicted value. Although it is desirable to apply a technique that yields a steeper word-specific slope of intelligibility, fragment normalisation also showed higher variability in its outcome, both in SRTn and slope. This technique also required higher average of level adjustments across with higher number of audio samples that needed to be capped at the limit of adjustment as described in the result of Study 1. Additionally, using the same reasoning behind Kollmeier's probabilistic model (Kollmeier (1990) as cited by (Zokoll et al., 2012)), adapting word-specific data with higher a standard deviation would reduce the overall steepness of list-specific slopes of intelligibility. Based on this argument, word normalisation would be the preferred method of normalisation for the MMST-AV as it would be more likely to produce higher list-specific slopes of intelligibility and better consistency in the test. Hence, word normalisation was used as the method of normalisation for all audio stimuli which were then used to generate sentences that were optimized for further evaluation and testing.

5.3.3 Study 3: Refinement of the Malay auditory-visual matrix sentence test

The third criteria for sentence selection was to identify the level at which judders between video frames were too noticeable to viewers. In this study, 10 participants were asked to view 2 sets of videos that contained 60 sentences each in auditory-visual mode. As described in Section 5.2.3.3 above, participants used a 10-point sliding scale from "no noticeable judder" at 0 to "highly noticeable judder" at 10. The results of this scoring in two trials are shown in Table 28 below.

Table 28: Table showing mean, median and standard deviation of participant's rating of the noticibility of judderin two trials.

Judder label	Trial 1			Trial 2		
	Mean	Median	StDev	Mean	Median	StDev
No judder	0.70	0.42	0.64	0.59	0.64	0.32
J1Tr01Ti1	1.63	1.29	1.66	0.87	0.37	1.06
J1Tr01Ti2	2.01	1.03	2.34	2.85	2.58	2.95
J1Tr01Ti3	6.00	5.97	1.34	5.58	6.24	2.83
J1Tr01Ti4	5.44	5.23	2.07	5.99	6.64	3.15
J1Tr02Ti1	2.80	2.07	2.92	1.19	0.39	1.48
J1Tr02Ti2	3.04	3.25	1.97	3.12	3.00	2.13
J1Tr02Ti3	8.20	8.51	1.84	6.43	7.62	3.63
J1Tr02Ti4	8.53	8.98	1.70	7.54	9.09	3.49
J1Tr03Ti1	1.93	1.33	2.01	1.21	0.32	1.51
J1Tr03Ti2	3.47	3.38	1.72	2.21	2.16	1.63
J1Tr03Ti3	5.81	6.14	2.38	4.78	4.81	3.27
J1Tr03Ti4	5.93	5.44	2.56	4.38	3.49	2.96
J2Tr12Ti1	1.18	0.85	1.05	0.80	0.36	0.95
J2Tr12Ti2	5.19	5.00	2.29	6.18	6.05	2.27
J2Tr12Ti3	8.01	7.91	1.62	7.76	8.40	2.62

J2Tr12Ti4	8.64	8.33	0.88	8.04	8.29	1.87
J2Tr13Ti1	0.45	0.12	0.72	0.74	0.27	1.08
J2Tr13Ti2	2.52	2.12	2.65	2.30	2.11	2.03
J2Tr13Ti4	8.21	8.00	1.66	8.10	8.53	1.48
J2Tr23Ti1	1.47	1.28	1.50	1.56	0.53	1.84
J2Tr23Ti2	3.58	2.99	2.66	3.57	3.09	2.11
J2Tr23Ti3	8.08	9.04	2.25	8.81	9.06	1.27
J2Tr23Ti4	8.65	8.82	1.44	8.87	9.41	1.27
Average	4.64	4.48	1.83	4.31	4.31	2.05

Average scores for Trial 1 and Trial 2 were 4.48 ± 3.15 and 4.3 ± 3.41 respectively. Using a paired sample t-test, no significant differences were observable between the two test trials (t(23)=0.93, p = 0.361), hence the two data were pooled together to provide a sum total sample size of twenty (n=20).

The distribution of scores for all judder labels according to rank is as shown in the figure below. A clear pattern can be seen where the original 100 continuous sentences (No judder) and sentences with tier 1 judder/s were rated least noticeable compared to sentences that contained tier 2, 3 or 4 judders.

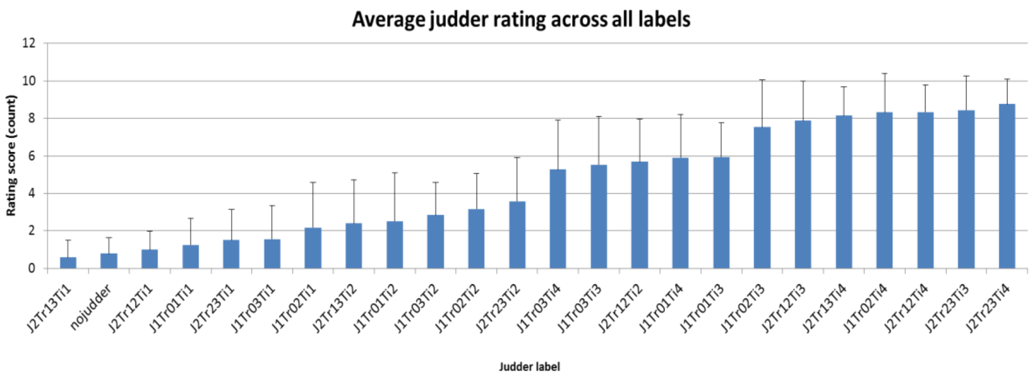


Figure 58: Distribution of average rating scores for all judder labels.

The mean rating score per sentence was 3.6 ± 2.92. As shown in Figure 59 below, a significant relationship was found between average rating scores and average judder pixel difference for each sentence using Pearson's correlation ($r = 0.921$, $n = 60$, $p < 0.00$).

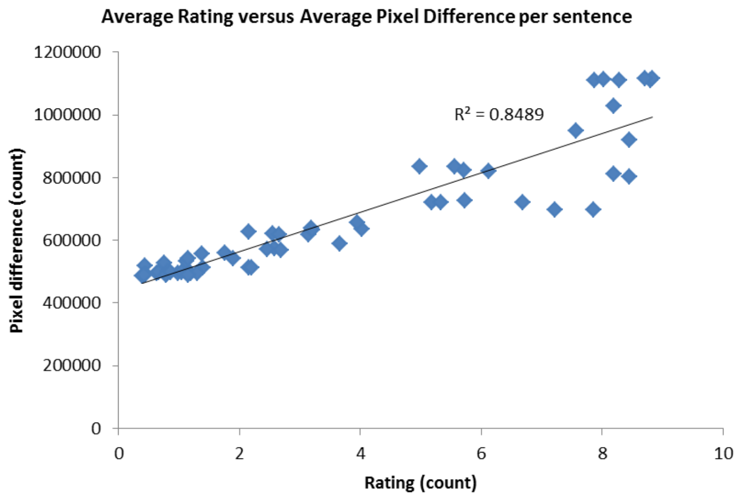


Figure 59: Scatterplot showing relationship between average participants' scores per sentence and average pixel difference per sentence.

To establish if a significant effect of judder type was present within sentences, a repeated measures analysis of variance (RM-ANOVA) was computed. It revealed a significant judder label effect was observable in the test, $F(81, 52.11) = 57.02$, $p < 0.01$ with Greenhouse-Geisser correction.

Table 29 below shows pairwise comparisons with Bonferroni correction between all judder labels. Significant differences were observable between all sentences with tier 2, 3 and 4 judders and the no judder sentences (except for J2Tr13Ti2, which contained Tier 2 judders at transitions 1 and 3).

Table 29: Pairwise comparison of judder ratings (judder labels compared to no judder) with Bonferroni correction.

Factor (Judder)		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
No judder	J1Tr01Ti1	-0.459	0.310	1.000	-1.896	0.978
	J1Tr01Ti2	-1.728	0.548	1.000	-4.265	0.810
	J1Tr01Ti3	-5.144*	0.435	0.000	-7.161	-3.128
	J1Tr01Ti4	-5.122*	0.548	0.000	-7.662	-2.582
	J1Tr02Ti1	-1.384	0.524	1.000	-3.813	1.044
	J1Tr02Ti2	-2.376*	0.413	0.004	-4.290	-0.463
	J1Tr02Ti3	-6.740*	0.591	0.000	-9.479	-4.000
	J1Tr02Ti4	-7.531*	0.511	0.000	-9.899	-5.164
	J1Tr03Ti1	-0.776	0.413	1.000	-2.688	1.135
	J1Tr03Ti2	-2.052*	0.423	0.031	-4.012	-0.091
	J1Tr03Ti3	-4.742*	0.572	0.000	-7.393	-2.091
	J1Tr03Ti4	-4.484*	0.580	0.000	-7.170	-1.797
	J2Tr12Ti1	-0.197	0.274	1.000	-1.466	1.072
	J2Tr12Ti2	-4.895*	0.507	0.000	-7.242	-2.548
	J2Tr12Ti3	-7.088*	0.524	0.000	-9.516	-4.660
	J2Tr12Ti4	-7.546*	0.398	0.000	-9.389	-5.704
	J2Tr13Ti1	0.200	0.065	1.000	-0.101	0.502
	J2Tr13Ti2	-1.615	0.461	0.653	-3.750	0.520
	J2Tr13Ti4	-7.362*	0.410	0.000	-9.264	-5.460
	J2Tr23Ti1	-0.725	0.348	1.000	-2.338	0.888
	J2Tr23Ti2	-2.783*	0.566	0.027	-5.406	-0.159
	J2Tr23Ti3	-7.652*	0.483	0.000	-9.890	-5.414
	J2Tr23Ti4	-7.970*	0.360	0.000	-9.637	-6.304

5.3.4 Generating normalised and refined auditory-visual test lists for the MMST-AV

Based on the three studies above, lists of sentences were generated for use in the MMST-AV. The sentence selections and list generation was based on the following guidelines:

i. All audio stimuli to be normalised using the word normalisation method in both TSN and BN.
ii. The highest amount of judder allowable for each sentence is tier 1, with no more than two judders per sentence.
iii. No sentence is to be repeated within or between lists in both noises.
iv. 15 lists with equal average predicted intelligibility score are to be generated, each containing 30 sentences. This number of sentences was chosen to ensure steepness of slope could also be reliably measured (Brand & Kollmeier, 2002).
v. Lists in the TSN should contain an equal frequency of words compared to the lists for the test in BN.
vi. Sentences should contain relatively equal occurences of phonemes between lists.

5.3.4.1 Normalizing audio stimuli using the word normalisation method (criterion i)

The average SRTn and slope for the MMST-AV in both noises are listed in the table below.

Table 30: Pre-normalisation and predicted post normalisation values for both TSN and BN.

	Pre-SRTn (dB SNR)	Pre-Slope (%/dB)	Post-SRTn (dB SNR)	Post Predicted Slope (%/dB)
Test specific noise	-9.03 ± 2.28	13.4	-8.95 ± 0.19	18.5
6-talker babble noise	-4.65 ± 1.83	11.2	-4.65 ± 0.2	14.1

Level adjustment decreased the overall spread of predicted responses and increased the predicted slope of intelligibility in both types of noise. The figures below show the normalisation process before and after level adjustment for individual words (SRTn and slope of intelligibility) and word categories (SRTn) for both types of noise.

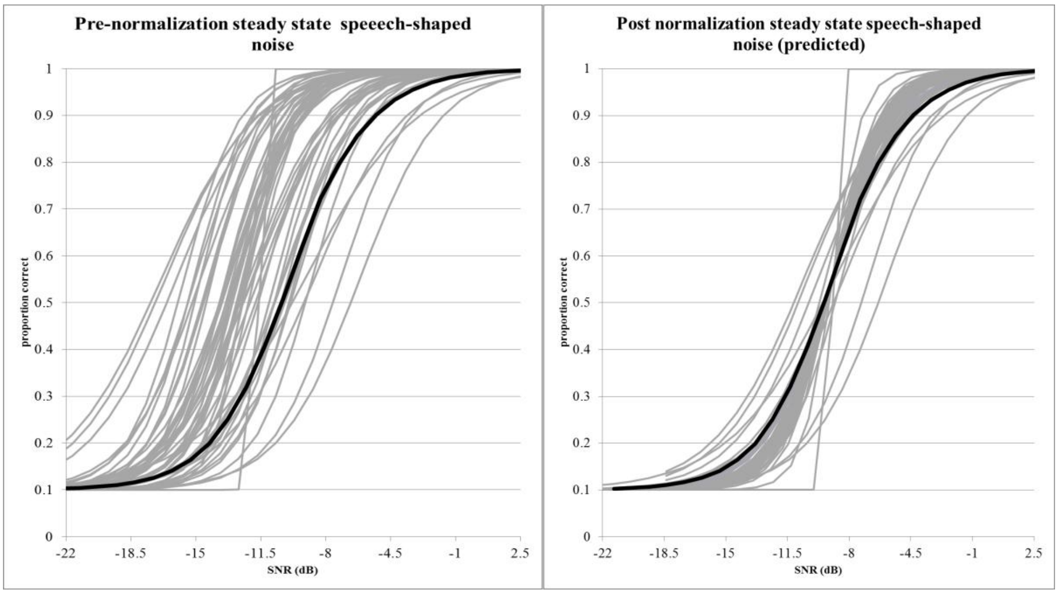


Figure 60: Psychometric functions of fifty disyllabic Malay words in test specific noise before normalisation (left) and the predicted functions after normalisation (right). Black line indicates the mean psychometric function.

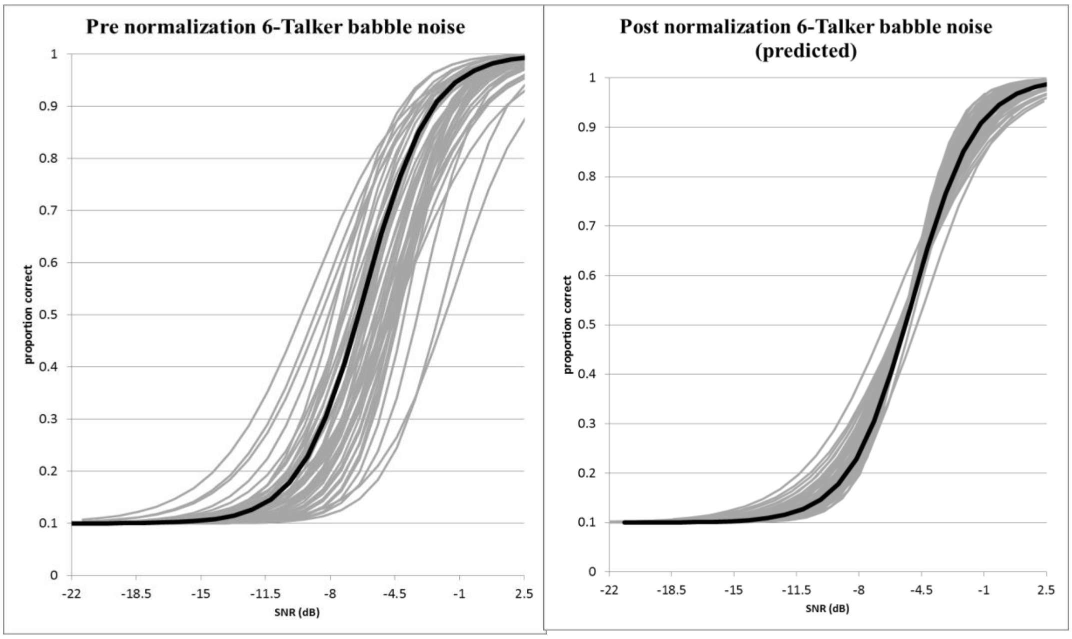


Figure 61: Psychometric functions of fifty disyllabic Malay words in six speaker babble noise before normalisation (left) and the predicted functions after normalisation (right). Black line indicates the mean psychometric function.

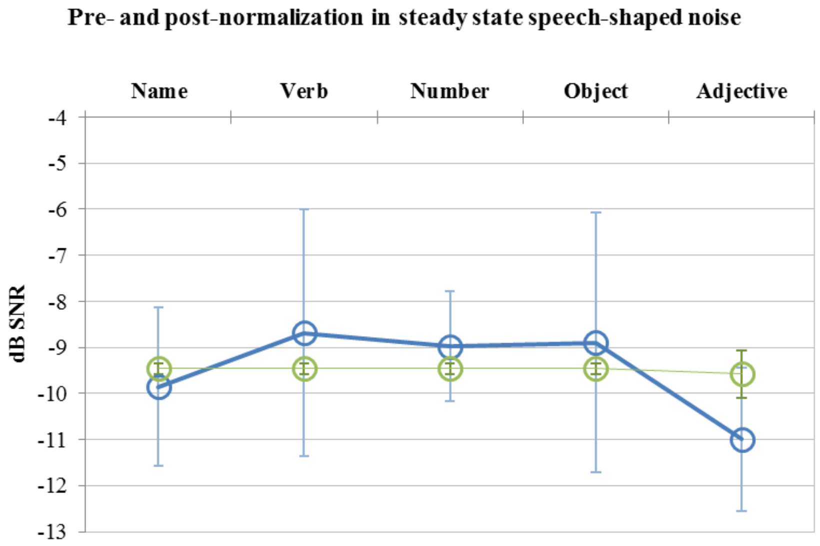


Figure 62: Word category mean SRTn values before (blue) and after (green) normalisation in test specific noise (TSN).

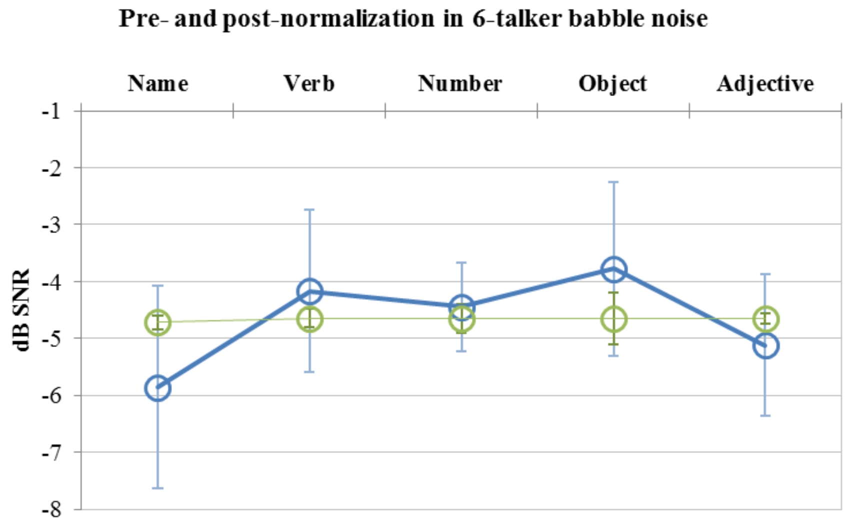


Figure 63: Word category mean SRTn values before (blue) and after (green) normalisation in 6-talker babble noise (BN).

5.3.4.2 Producing unique sentences with no judder or tier 1 judders for both TSN and BN (Criteria ii & iii).

Producing unique sentences allows for greater diversity of word distribution within lists and between the two tests in different types of noise. It also eliminates any chance of memorizing the sentences as well as using the full potential of the matrix sentence test structure. To create 15 lists with 30 sentences each, a total of 1200 unique sentences were generated. The sentences only contained either Tier 1 judders and/or the no judder continuous video frames from the original recording of 100 continuous sentences.

The initial step was to eliminate all possible sentence combinations that could include judder Tiers 2 to 4, which was done using a custom VI written by Assoc. Prof. Greg O'Beirne. All possible word combinations were then iteratively shuffled based on their individual word intelligibility score to form groups of 15 sentences that had almost equal predicted list-specific intelligibility scores.

Summary of the generated list slope distribution is shown in Tables 30 and 31 below.

Table 31: Distribution of predicted list-specific slope of intelligibility in TSN.

Test list in TSN	**Mean Slope**	**StDev**	**Max**	**Min**
List 1	15.45%/dB	1.63%/dB	19.13%/dB	12.94%/dB
List 2	15.39%/dB	1.16%/dB	17.27%/dB	13.10%/dB
List 3	15.22%/dB	1.45%/dB	18.06%/dB	12.38%/dB
List 4	15.49%/dB	1.34%/dB	18.20%/dB	13.27%/dB
List 5	15.29%/dB	1.41%/dB	18.83%/dB	12.89%/dB
List 6	15.93%/dB	1.94%/dB	19.43%/dB	12.99%/dB
List 7	15.54%/dB	1.20%/dB	18.00%/dB	13.39%/dB
List 8	15.68%/dB	2.16%/dB	20.57%/dB	11.38%/dB
List 9	15.76%/dB	2.06%/dB	20.45%/dB	12.21%/dB
List 10	15.54%/dB	1.36%/dB	18.56%/dB	13.67%/dB
List 11	15.25%/dB	1.40%/dB	19.76%/dB	13.03%/dB
List 12	15.50%/dB	1.44%/dB	18.43%/dB	12.64%/dB

Test list in TSN	Mean Slope	StDev	Max	Min
List 13	15.23%/dB	1.28%/dB	18.07%/dB	13.40%/dB
List 14	15.11%/dB	1.57%/dB	17.55%/dB	11.58%/dB
List 15	14.93%/dB	1.15%/dB	16.83%/dB	13.17%/dB
Average	15.43%/dB	1.50%/dB	18.61%/dB	12.80%/dB

Table 32: Distribution of list-specific slope of intelligibility in BN.

Test list in BN	Mean Slope	StDev	Max	Min
List 1	13.93%/dB	0.76%/dB	15.46%/dB	12.38%/dB
List 2	13.97%/dB	0.79%/dB	16.32%/dB	12.82%/dB
List 3	13.94%/dB	0.91%/dB	16.05%/dB	12.30%/dB
List 4	13.99%/dB	0.79%/dB	15.46%/dB	12.39%/dB
List 5	14.08%/dB	0.82%/dB	15.48%/dB	12.46%/dB
List 6	14.00%/dB	1.00%/dB	15.84%/dB	12.10%/dB
List 7	13.93%/dB	0.90%/dB	15.73%/dB	11.99%/dB
List 8	13.86%/dB	1.21%/dB	16.21%/dB	12.43%/dB
List 9	13.87%/dB	0.71%/dB	14.81%/dB	12.09%/dB
List 10	13.90%/dB	1.11%/dB	16.50%/dB	12.09%/dB
List 11	13.68%/dB	0.92%/dB	15.40%/dB	12.19%/dB
List 12	14.02%/dB	1.01%/dB	15.61%/dB	12.32%/dB
List 13	14.10%/dB	0.71%/dB	15.58%/dB	13.11%/dB
List 14	14.16%/dB	0.84%/dB	15.74%/dB	12.65%/dB
List 15	14.04%/dB	1.04%/dB	16.09%/dB	12.47%/dB
Average	13.96%/dB	0.90%/dB	15.75%/dB	12.38%/dB

5.3.4.3 Comparison of word and phoneme distribution within MMST-AV lists (Criteria iv to vi)

The distribution of words selected between tests was similar, as shown in Table 33 below.

Table 33: Distribution of selected words for each test lists in TSN and BN

Constant	**Count**	**Babble**	**Count**	**Word difference (constant - babble)**	
abang	71	abang	67	abang	4
ada	63	ada	59	ada	4
adik	60	adik	55	adik	5
ambil	97	ambil	83	ambil	14
ayah	66	ayah	74	ayah	-8
bagi	17	bagi	19	bagi	-2
baju	57	baju	48	baju	9
banyak	60	banyak	47	banyak	13
baru	52	baru	45	baru	7
beri	54	beri	44	beri	10
besar	57	besar	59	besar	-2
bola	17	bola	19	bola	-2
buku	63	buku	59	buku	4
cantik	63	cantik	61	cantik	2
dapat	55	dapat	50	dapat	5
dia	57	dia	48	dia	9
dua	55	dua	58	dua	-3
empat	65	empat	79	empat	-14
enam	66	enam	51	enam	15
hijau	61	hijau	52	hijau	9
hitam	67	hitam	78	hitam	-11
ibu	56	ibu	62	ibu	-6
kakak	63	kakak	79	kakak	-16
kami	55	kami	54	kami	1
kecil	54	kecil	55	kecil	-1
kita	63	kita	52	kita	11

Constant	**Count**
kotak	90
kunci	41
lama	60
lampu	36
lapan	58
lima	60
mahal	58
mahu	37
mangkuk	104
meja	58
merah	65
minta	54
nampak	100
nenek	62
perlu	33
pisau	79
putih	63
satu	53
saya	47
semua	56
suka	90
tiga	70
topi	55
tujuh	57
Grand Total	3000

Babble	**Count**
kotak	87
kunci	48
lama	46
lampu	46
lapan	65
lima	67
mahal	84
mahu	48
mangkuk	98
meja	47
merah	54
minta	66
nampak	105
nenek	69
perlu	41
pisau	80
putih	66
satu	52
saya	40
semua	50
suka	85
tiga	46
topi	68
tujuh	85
Grand Total	3000

Word difference (constant - babble)	
kotak	3
kunci	-7
lama	14
lampu	-10
lapan	-7
lima	-7
mahal	-26
mahu	-11
mangkuk	6
meja	11
merah	11
minta	-12
nampak	-5
nenek	-7
perlu	-8
pisau	-1
putih	-3
satu	1
saya	7
semua	6
suka	5
tiga	24
topi	-13
tujuh	-28
Grand Total	0

.

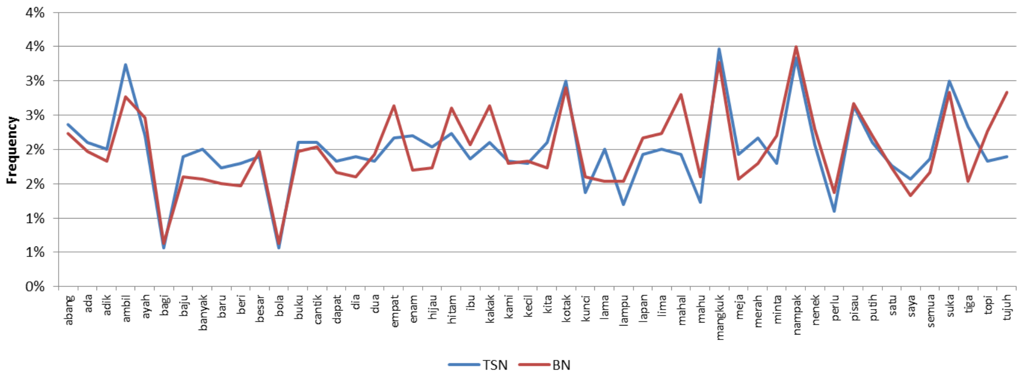


Figure 64: Comparison between word distribution between lists used in test specific noise (TSN) and 6-talker babble noise (BN)

The distribution of words between lists in the TSN and BN were not equal. Some lists had no words represented and up to 6 same word repetitions per list. This is because sentences were generated by manually shuffling words to achieve optimum and equal average list-specific slope scores and considerations were made to ensure at least only Tier 1 judders were present. Shown below is the word distribution between and within lists.

Table 34: Distribution of words for test lists in TSN.

TSN	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10	List 11	List 12	List 13	List 14	List 15
abang	3	4	4	1	3	3	1	4	5	1	4	4	5	3	4
ada	4	1	3	4	3	3	5	2	5	1	2	5	4	2	4
adik	3	6	1	6	4	4	1	3	3	2	5	1	2	3	1
ambil	5	5	6	4	4	3	4	2	6	6	3	6	7	3	6
ayah	3	1	4	3	5	4	5	2	3	5	2	5	2	4	4
bagi	1	1	1	0	3	0	1	1	0	1	0	0	0	0	0
baju	3	6	2	3	5	2	3	5	4	1	1	2	2	5	2
banyak	3	1	4	2	1	7	0	5	5	5	1	5	3	1	0
baru	2	3	3	5	4	3	6	3	2	2	3	2	1	4	2
beri	1	0	3	3	4	4	2	3	3	6	4	2	1	3	3
besar	1	4	2	1	5	2	2	4	4	3	2	3	3	0	3
bola	1	1	1	0	3	0	1	1	0	1	0	0	0	0	0
buku	4	1	3	4	3	3	5	2	5	1	2	5	4	2	4
cantik	2	3	2	3	4	3	4	2	1	6	2	2	4	5	7
dapat	3	5	2	3	5	2	3	5	3	2	1	2	1	5	3

TSN	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10	List 11	List 12	List 13	List 14	List 15
dia	6	3	3	4	4	3	3	2	2	2	2	3	4	1	3
dua	1	3	5	2	5	4	3	1	2	2	2	2	5	2	3
empat	3	5	6	5	0	1	3	3	3	3	4	3	3	3	5
enam	1	4	2	3	5	5	3	3	5	3	5	3	5	3	4
hijau	1	4	2	3	2	6	2	4	4	2	5	1	4	3	2
hitam	1	6	5	5	1	2	2	3	6	2	6	5	4	4	1
ibu	2	2	0	5	0	3	4	1	5	1	2	3	1	5	6
kakak	7	4	4	0	1	2	6	4	2	3	5	5	6	2	2
kami	3	2	3	2	4	1	3	2	3	4	2	1	1	5	2
kecil	2	1	3	2	3	0	3	3	2	3	5	2	1	4	4
kita	1	3	3	1	5	1	3	5	3	5	3	3	2	2	2
kotak	4	6	5	4	3	2	3	2	5	6	3	5	7	4	5
kunci	4	3	2	1	2	1	1	4	3	0	2	0	3	4	0
lama	5	1	4	3	4	2	2	3	4	4	2	4	3	1	2
lampu	0	1	2	1	3	1	2	0	2	2	0	2	3	2	7
lapan	3	2	3	2	4	1	5	3	1	4	3	1	3	5	6
lima	5	2	1	3	3	4	1	0	5	1	2	4	3	7	3
mahal	9	1	3	1	0	2	3	4	2	3	3	2	6	1	1
mahu	4	4	1	2	1	2	2	3	1	1	1	1	2	3	0
mangkuk	5	4	6	7	3	8	7	5	5	5	7	6	4	7	4
meja	1	0	3	3	4	4	3	3	2	6	5	3	1	2	4
merah	3	5	2	5	3	5	1	2	3	2	1	5	1	2	4
minta	4	3	1	4	1	2	1	5	0	4	6	6	4	2	1
nampak	4	5	6	5	3	9	6	3	5	4	8	3	4	8	4
nenek	1	4	5	3	3	7	4	3	1	4	3	1	3	3	4
perlu	0	2	1	1	3	1	1	0	2	1	0	2	4	1	5
pisau	4	4	5	3	2	6	4	4	3	6	4	3	2	2	3
putih	4	2	4	2	4	5	5	2	2	3	1	4	3	6	4
satu	2	3	0	5	4	4	7	3	3	2	3	2	3	5	2
saya	1	1	3	5	1	2	0	4	3	3	2	4	4	2	2
semua	3	4	3	2	4	1	3	6	2	2	4	4	0	1	3
suka	4	4	6	4	3	4	5	6	5	4	5	3	3	3	4
tiga	3	5	2	3	4	2	3	3	2	5	3	5	2	0	3
topi	4	4	1	4	2	3	1	4	1	2	6	4	4	2	1
tujuh	6	1	4	3	0	1	2	3	2	3	3	1	3	3	1
	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150

Table 35: Distribution of words for test list in BN.

BN	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10	List 11	List 12	List 13	List 14	List 15
abang	3	4	2	2	4	5	3	4	1	5	3	2	3	5	4
ada	2	4	2	3	3	3	5	4	3	2	5	3	3	1	6
adik	1	3	4	2	3	3	3	2	3	5	1	1	4	1	4
ambil	5	5	4	3	5	3	3	3	4	3	4	5	2	3	8
ayah	6	8	1	3	8	5	4	2	2	1	5	6	4	6	1
bagi	1	1	0	2	1	0	1	1	1	2	0	0	0	1	2
baju	3	3	3	4	2	2	3	3	4	3	3	0	1	4	0
banyak	3	1	5	2	0	4	5	2	3	0	3	2	4	2	1
baru	1	1	2	3	2	1	2	1	1	1	2	2	5	2	3
beri	3	2	1	2	2	2	3	3	2	1	1	1	1	2	4
besar	3	1	2	3	3	1	2	3	1	4	5	2	2	5	4
bola	1	1	0	2	1	0	1	1	1	2	0	0	0	1	2
buku	2	4	2	3	3	3	5	4	3	2	5	3	3	1	6
cantik	2	2	5	2	4	1	2	2	4	2	2	5	5	4	4
dapat	2	3	3	4	2	2	3	2	4	4	3	1	1	5	0
dia	1	2	3	1	0	3	1	3	2	2	3	1	3	1	3
dua	1	2	3	2	3	2	2	2	1	2	4	2	4	5	4
empat	5	5	4	6	6	5	2	4	5	4	3	4	5	3	3
enam	1	3	2	6	2	1	3	4	4	4	1	3	0	4	3
hijau	2	3	5	5	1	2	5	2	2	6	0	1	1	3	0
hitam	3	3	3	2	6	7	2	4	4	4	2	3	4	4	6
ibu	7	4	3	1	2	1	4	3	6	2	5	6	2	2	0
kakak	3	1	7	7	1	4	3	3	5	4	1	3	3	3	11
kami	2	2	1	3	3	3	4	7	5	1	2	0	3	3	3
kecil	3	3	2	1	4	5	2	6	3	3	2	0	2	2	2
kita	4	4	1	3	4	1	2	2	0	5	4	2	0	4	3
kotak	6	4	4	4	6	3	3	4	4	3	5	5	3	3	9
kunci	2	3	4	2	2	2	2	1	1	3	3	3	3	0	1
lama	2	1	3	1	0	1	5	2	3	1	3	2	3	2	2
lampu	1	2	3	2	3	4	2	4	3	3	3	2	3	2	1
lapan	3	2	3	4	5	1	3	2	3	1	2	6	2	5	5
lima	4	3	3	2	5	4	1	3	3	3	5	4	2	6	2
mahal	3	7	4	8	2	6	6	4	5	2	4	4	5	2	6
mahu	4	1	4	2	3	2	2	1	2	3	2	2	4	0	1
mangkuk	3	6	4	6	2	8	5	3	3	6	4	1	5	8	3
meja	2	2	1	3	2	2	3	3	2	1	1	1	1	4	5

BN	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10	List 11	List 12	List 13	List 14	List 15
merah	4	2	2	2	1	3	2	4	5	4	5	5	0	2	2
minta	1	4	3	3	3	4	4	1	5	4	3	6	4	2	2
nampak	3	6	5	7	4	8	6	5	3	6	6	2	5	8	4
nenek	2	1	5	4	5	2	3	2	5	5	3	5	5	2	1
perlu	1	2	2	2	2	4	2	4	3	2	3	1	3	2	1
pisau	8	1	6	2	4	2	2	6	3	3	3	8	7	6	1
putih	7	7	2	3	7	3	2	2	2	3	5	6	3	4	1
satu	2	2	1	1	2	1	4	3	2	2	0	2	6	1	3
saya	1	1	3	4	0	3	3	2	1	0	3	4	3	3	0
semua	3	3	3	1	2	4	0	2	4	5	2	2	2	0	3
suka	8	2	6	2	5	2	1	6	3	3	3	9	7	6	2
tiga	4	3	3	2	1	3	1	3	2	6	4	1	0	2	2
topi	2	4	3	2	5	4	4	1	6	4	3	7	4	1	2
tujuh	4	6	3	4	4	5	9	5	3	3	6	4	5	2	4
	150	150	150	150	150	150	150	150	150	150	150	150	150	150	150

The maximum number of times a word appeared in a list of 30 sentences was 9 (i.e. 30% of choices for a particular column), which occurred for mahal and nampak in TSN and kotak, suka, and tujuh in BN. However phoneme distribution within and between test list was similar with all phoneme represented, as shown in Figures 65 and 66 below.

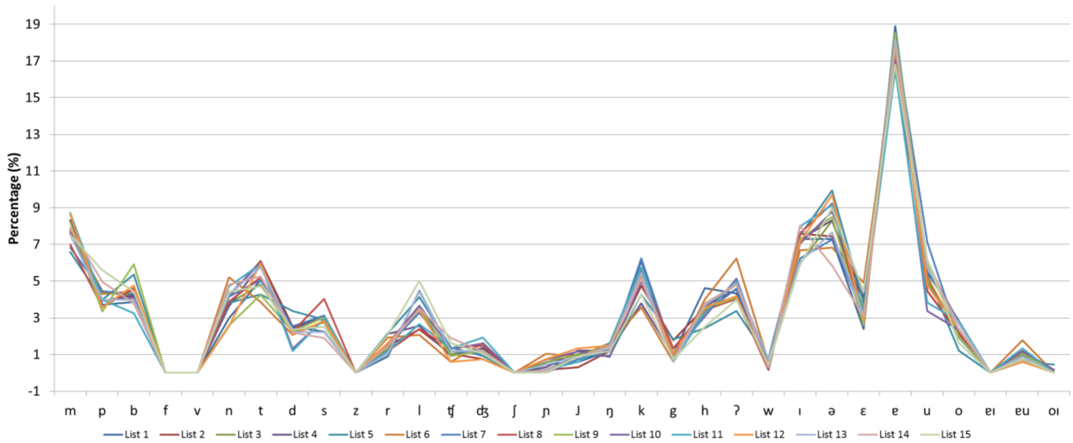


Figure 65: Phoneme distribution of lists used in test specific noise (TSN).

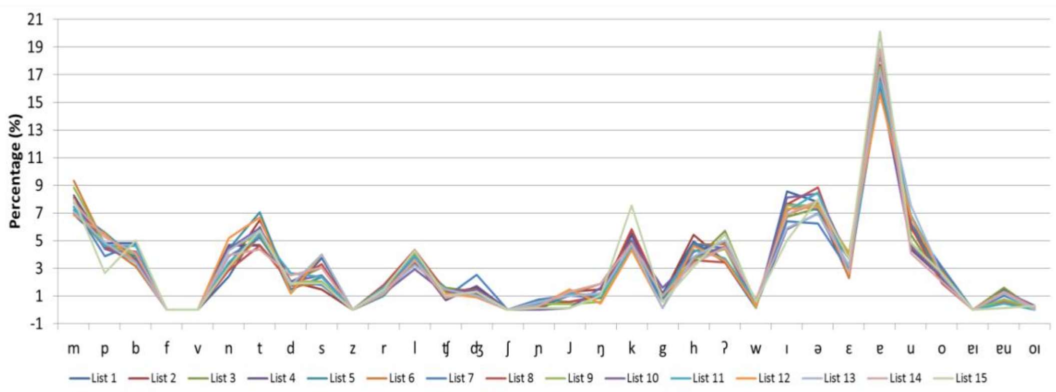


Figure 66: Phoneme distribution of lists used in 6-talker babble noise (BN).

5.4 Discussion

5.4.1 Normalisation and refinement of MMST-AV.

In this study, normalisation processes were conducted concurrently for both test specific noise and 6-talker babble noise in Malaysia using different sets of normal-hearing listeners. This was due to the long duration of testing and the limited time available to collect this data in Malaysia. The aim of normalizing the intelligibility of words and fragments in a particular type of noise was done by level adjusting the SRTn of each word or fragment to match the overall average SRTn between the two methods of normalisation within that noise. This was achieved in these two normalisation methods with a limit on level adjustment of ±3 dB to ensure no abnormal loudness fluctuations were present within sentences. Normalisation increased the predicted slope by an average of 4.9%/dB in the test specific noise and by an average of 2.8%/dB in the 6-talker babble noise. These increases in slope are consistent with other predicted increases of slope in other MST tests (5.1%/dB in the Spanish MST (Hochmuth et al., 2012) and 4.5%/dB in the Danish MST (Wagener et al., 2003), for example). Using test specific noise also produced steeper pre-normalisation and predicted post-normalisation slopes compared to the 6-talker babble noise. As expected, due to the informational masking and level modulation provided by the 6-talker babble noise caused the average SRTn to be higher than in the test specific noise. As there are competing number of speakers in BN, the sensitivity of the test was reduced as listeners were tasked to resolve the speech stimuli in a speech-on-speech masking scenario, which is known to increase attentional load and reduce performance in responses (Van Engen et al., 2014). As shown in Table 30 above, the SRTn for babble noise was higher by an average of 4.3 dB in both normalisation methods, which is consistent with previous findings of other MSTs using multiple talker babble noise (Hochmuth, Jürgens, et al., 2015; Wagener & Brand, 2005). 15 lists were generated containing 30 unique sentences for both types of noise. The level adjustments determined in this study were applied and then used to determine the optimal normalisation method for the MMST-AV in Study 2.

Evaluation of the optimal normalisation method was conducted on nine normal hearing Malay speaking adults in Christchurch, New Zealand. Using the scoring method proposed in Trounson (2012) to identify word-specific scores in both normalisation methods, it was found that there were no statistical differences seen in the SRTn levels between both word and fragment normalisation. However, a difference in slope was observable between the two types of normalisation, with the fragment normalisation showing slope of intelligibility that was steeper by 1.5%/dB. Using fragment normalisation, the range of responses obtained showed large spread of data, hence producing higher standard deviations in both SRTn and slope scores. According to the probabilistic model by Kollmeier (1990) in Zokoll et al. (2012), a predicted slope of 3.5%/dB should be expected from using the outcome of fragment normalisation, compared to a predicted slope of 11.5%/dB if the data from word normalisation were used. Normalising based on the individual audio fragments could have caused non-natural amplitude changes within words, making it more confusing for listeners. This could explain the large variation in percentage scores between listeners when tested at fixed SNRs. Inconsistency was also observable in the larger variation in the SRTn and slope measurements using this method. A more consistent and predictable outcome is desired, and so the word normalisation method was selected and used as the standard normalisation method for the evaluation and validation of the MMST-AV.

The third part of the study involved investigated the noticeable movements between video frames that could make the video look unnatural. Consistency throughout video recording is essential, as the coupling between the visual and audio signals is important for viewers to: (a) form meaning of the speech (Campanella & Belin, 2007); (b) avoid confusion by mismatched timing of both modalities (Summerfield, 1992); and (c) avoid phoneme confusion, as seen in the McGurk's effect studies (Jones & Callan, 2003; Macdonald & McGurk, 1978). This part of the study showed that the judder levels in the original 100 continuous recording were not different from the measured judder levels that were within one standard deviation of the mean of those original levels, (termed Tier 1 for this study). This was true even when two Tier 1 judders were present within the same sentence, suggesting the consistency in rating of participants. In one exception, the "noticeability" ratings of sentences containing Tier 2 judders at position 1 and 3 were found to not be statistically different to the "no judder" sentences. To avoid any confusion, the judder label J2Tr13Ti2 was still excluded from the final version of the test. Exclusions of Tiers 2, 3 & 4 from the test lead to the elimination of half of the total possible sentence manifestations (> 50 000 sentences), which had a direct effect on the number of possible unique sentences that could be produced for the final version of the test.

5.4.2 Generating normalised and refined auditory-visual test lists for the MMST-AV

As several rules were created to ensure the audio stimuli were normalised and the video stimuli were refined, the number of possible sentences available for the final test was reduced, which caused the word distributions between lists and between test noise types to not be balanced. Poor uniformity and the possible common recurrence or absence of certain words could affect overall sensitivity of the test (Houben & Dreschler, 2015). However, based on Hochmuth, Jürgens, et al. (2015) the test format and language used had no influence on listeners' performance in various versions of the MST. However, they were unable to conclude whether specific word combinations could have a direct effect on listeners, as the test involves a relatively small amount of speech material. Even so, the phonemic composition of the tests in the current study was examined and found to be relatively equal within lists in each noise type. It is expected that this will not affect the homogeneity of the test lists even when the distribution of words is random, because the sum average of lists produced similar predicted list-specific intelligibility scores.

In the coming chapter we examine if the disproportionate distribution of words had any effect on listeners' performance in both SRTn and slope.

CHAPTER 6

EVALUATION OF LIST EQUIVALENCY IN THE MALAY DIGIT TRIPLET AND MATRIX SENTENCE TESTS

6.1 Evaluation of the Malay Digit Triplet Test

6.1.1 Introduction

Evaluation is the process of verifying the normalisation process. This is commonly done by obtaining responses from normal hearing listeners at either 2 or 3 fixed signal-to-noise ratios, and fitting an intelligibility function to them using nonlinear regression. This process does not only provide verification of the normalisation process by looking for consistency between lists but also provides experimenters with the normative values for normal hearing listeners. For other versions of digit triplet tests, the SRTn values in fixed (i.e. non-adaptive) measurement using headphones have been reported to be between as low as -10.5 dB SNR in the French DTT (Jansen et al., 2010), to as high as -6.9 dB SNR for the Swedish DTT shown in Table 3 of Zokoll et al. (2012). Slopes of intelligibility have ranged from 24.2%/dB in the Swedish DTT to 16% in the Dutch DTT. Using telephones typically reduced the overall sensitivity by 2 to 4 dB. The SRTn using telephones was between -7.5 dB SNR for the Polish DTT, (Ozimek, Kutzner, Sęk, et al., 2009) and -4.3 dB SNR for the Swedish DTT, shown in Table 3 of Zokoll et al. (2012). The slopes using telephones ranged from 19.2%/dB to 24.3%/dB. As described above, the results of multilanguage DTT versions are varied, and it is expected that the Malay version of the DTT test should yield results close to the range described above. The aim of this study was evaluate the Malay DTT: (a) with fixed and adaptive measurements; (b) using headphones and telephones; (c) in test specific noise (test specific noise, or TSN) and spectrotemporal gap noise (STG).

6.1.2 Methods

6.1.2.1 Recruitment of participants

6.1.2.1.1 Sample size calculation

For the evaluation process of the Malay digit triplet test, an automated sample size calculator estimated that at least 12 hearing impaired subjects had to be recruited for this study to be able to reject the null hypothesis (Dupont & Plummer, 1990). This was to achieve a 90% power of study, which is based on the response of the impaired hearing group in Smits et al. (2004) with the standard deviation of 0.3 dB when the noticeable difference of the test is 1 dB.

Participants were divided into two groups: Group 1 was scheduled to complete the task using fixed SNR measurements; and Group 2 was given the task of completing the evaluation using adaptive measurement.

Group 1 consisted of 16 adult Malay native speakers aged 22.3 ± 1 year old (6 male and 10 female). All participants had normal hearing (hearing threshold average < 20 dB HL at 500, 1000. 2000, 4000 & 8000 Hz). Participants were undergraduate students at the University of Canterbury, New Zealand. They were given NZD 20 petrol vouchers for their time and effort.

Group 2 consisted of 20 adult Malay native speakers aged 26.4 ± 6.9 years old (5 male and 15 female). All participants had normal hearing (hearing threshold average < 20 dB HL at 500, 1000. 2000, 4000 & 8000 Hz). Participants were either staff or undergraduate students at the International Islamic University, Malaysia. They were paid RM50 cash for their time and effort.

6.1.2.2 Test procedure I: Fixed SNR measurement

The evaluations using fixed SNR were all conducted in a single-walled audiometric cabin at Department of Communication Disorders research facility, Level 8 Rutherford building at the University of Canterbury, New Zealand. Prior to the evaluation process, hearing threshold levels were examined at frequencies 250, 500, 1000, 2000, 4000 & 8000 Hz. Group 1 participants were asked to sit in front of the computer and use the computer keyboard to select the digits they heard. Participants were tested in a sequence of preset lists to save time because testing time for all the lists for each participant would be too long and this could affect attention and performance. The pre-set list was designed so that each participant was only tested using half of the total available lists. Effectively, the sum data of all test lists is equivalent to only eight participants. The result of all participants for the two signal-to-noise ratios were averaged and used in the logistical function equation (2) to produce the SRTn and slope of intelligibility.

To present the test with normalised digit triplets, the UCAST platform software was used run on a Windows™ PC using an external sound card (Creative X-Fi 51 Soundblaster sound card). The transducers used for this test were Sennheiser HD 280 Pro headphones and a generic telephone receiver. The telephone receiver was coupled to the sound card via JK Audio THAT-2 audio handset tap. As the normalisation process was done separately for the type of transducers and background noise, specific SNR levels for each test condition were selected to approximate the 40% (SNR_1) and 80% score (SNR_2) of the intelligibility function. The SNR levels used for this test were chosen based on the average triplet scores for each condition. The SNR levels for each condition are listed below.

Table 36: Fixed level signal-to-noise ratio used for list equivalency measurement.

Test conditions	SNR_1 (dB SNR)	SNR_2 (dB SNR)
Headphone in TSN	-13.2	-10.4
Headphone in STG	-13.2	-10.4
Telephone in TSN	-12	-9
Telephone in STG	-12.2	-10

Background noise was presented monaurally at a constant 65 dB SPL throughout testing and the signal-to-noise ratios were adjusted by changing intensity levels of digit triplets.

6.1.2.3 Test procedure II: Adaptive measurement

Group 2 participants were all based in Malaysia and was tested in a double walled audiometric booth (average ambient noise = 22.5 dB SPL, reverberation time, R_{T60}= 0.1 millisecond). Participants were asked to key in their responses on a keyboard. Only the speech reception threshold in noise was identified using Equation 2 in an adaptive staircase procedure for both telephones and headphone application and in both types of background noise. The adaptive procedure involved recording and evaluating the response obtained at one level of presentation to determine the next level of presentation until the definition of target threshold has been achieved using a simple 1-up, 1-down staircase that tracked the 50% level (Levitt, 1970). For the Malay DTT, the initial presentation was set at 2 dB SNR and step size reduces and increases at a constant 2 dB after each presentation. An example of the adaptive staircase method for the MDTT is shown below.

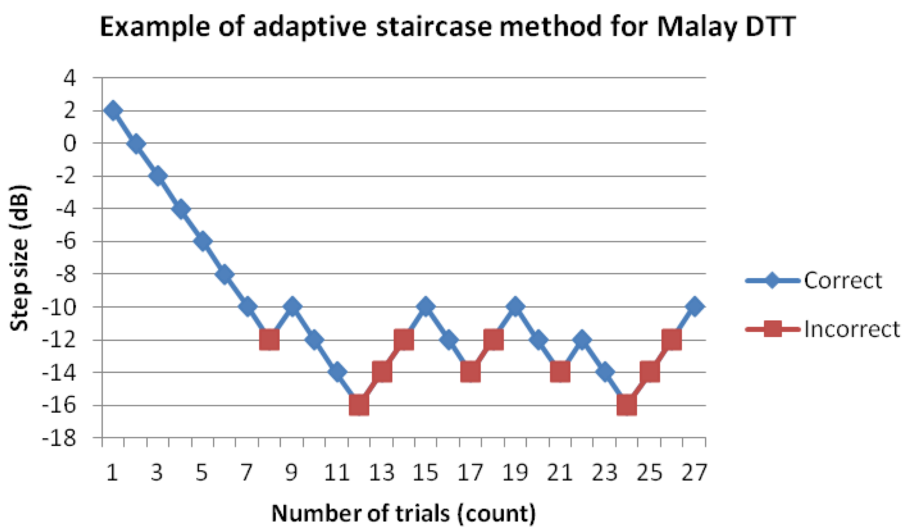


Figure 67: Adaptive staircase method with fixed step size in MDTT.

This method is widely used in other DTT tests and was based on the proposed methods in Plomp & Mimpen (1979). Plomp & Mimpen (1979) suggested that at least 13 trials are necessary to arrive at the target SRTn reliably; a revision of this method was done by Smits et al. (2004) and resulted in a proposed 20 triplet presentation to be the minimum amount of digit triplets necessary to reliably screen listeners with varying levels of hearing. For the MDTT, each list contained 27 unique digit triplets and all Group 2 participants completed all 8 lists to evaluate their equivalency.

6.1.3 Results and analyses

6.1.3.1 Fixed SNR measurement

Group 1 participants' average hearing threshold at octave frequencies 500 Hz to 8000 Hz were 6.9 dB HL ± 4.52. The average hearing thresholds across the audiometric range are shown the figure below.

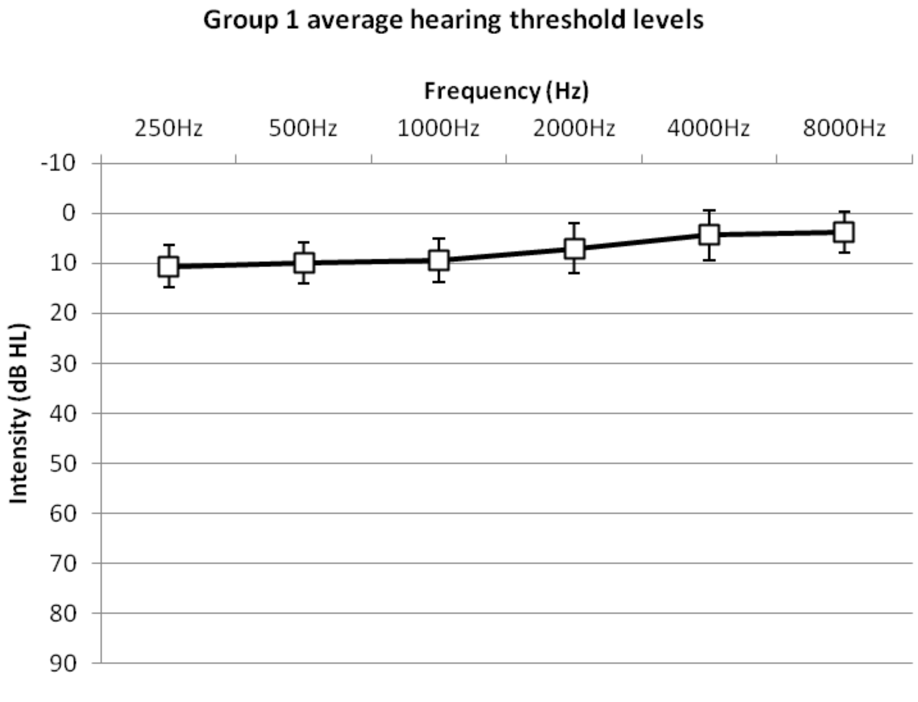


Figure 68: Group 1 average hearing threshold levels (n=16).

The distribution of SRTn and slope measured in fixed SNR using headphone in test specific noise are shown in Table 37 and Figure 69 below. For the list evaluation using headphone in TSN, the triplet list 7 showed the lowest average SRTn at -11.75 dB and the highest recorded SRTn was found in list 3 at -10.67 dB SNR. The steepest slope is seen for list 6 at 19.1 %/dB and the shallowest slope is seen in list 3 at 10.4 %/dB.

The approximated SNR level for 80% intelligibility score (SNR_2) used in all four test conditions produced lower than expected scores which is not ideal to ensure the accurate estimates for SRTn and slope. Nevertheless, the results showed good consistency throughout all four test conditions with average standard deviations of list SRTn score of less than 0.4 dB SNR.

Table 37: Average SRTn and slope for all 8 lists in the MDTT using headphones in TSN.

Headphones in TSN	SNR_1 (%)	SNR_2 (%)	SRT_n	Slope at midpoint:
MalayTripletList1	23%	64%	-11.32	15.93 %/dB
MalayTripletList2	21%	63%	-11.21	16.41 %/dB
MalayTripletList3	26%	53%	-10.67	10.37 %/dB
MalayTripletList4	22%	59%	-11.04	14.53 %/dB
MalayTripletList5	25%	71%	-11.66	18.08 %/dB
MalayTripletList6	19%	66%	-11.26	19.05 %/dB
MalayTripletList7	28%	70%	-11.75	15.98 %/dB
MalayTripletList8	21%	64%	-11.26	17.05 %/dB
Average	23%	64%	-11.27	15.92 %/dB
Standard deviation	3%	6%	0.34	2.64 %/dB

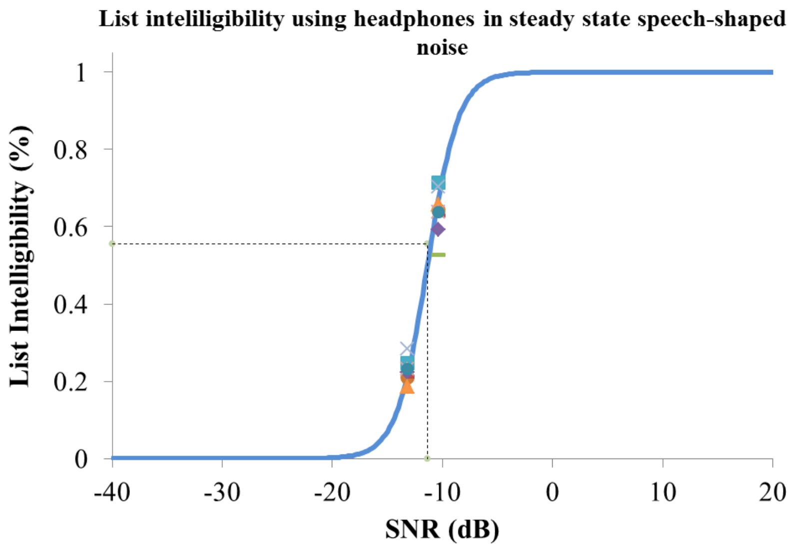


Figure 69: Measured performance-intensity function in TSN for all lists in the MDTT using headphones.

The average SRTn using headphones in STG was slightly lower than the average for headphones in TSN. However the average slope in this condition was shallower than the previous test condition at 14.2 %/dB compared to 15.9%/dB. Highest SRTn and slope scores were recorded in list 4 at -10.79 dB SNR and list 3 at 26.77%/dB respectively. The distribution of results measured in fixed SNR for SRTn and slope using headphone in the spectrotemporal gap noise are shown in Table 38 and Figure 70 below.

Table 38: Average SRTn and slope for all 8 lists in the MDTT using headphones in STG.

Headphones in STG	SNR_1 (%)	SNR_2 (%)	SRT_n	Slope at midpoint:
MalayTripletList1	31%	80%	-12.20	19.46 %/dB
MalayTripletList2	31%	63%	-11.88	15.59 %/dB
MalayTripletList3	19%	69%	-11.56	21.64 %/dB
MalayTripletList4	28%	60%	-11.25	12.22 %/dB
MalayTripletList5	41%	62%	-11.99	9.73 %/dB
MalayTripletList6	36%	65%	-11.83	10.67 %/dB
MalayTripletList7	36%	74%	-12.20	14.59 %/dB
MalayTripletList8	42%	74%	-12.54	12.26 %/dB
Average	33%	68%	-11.93	14.27 %/dB
Standard deviation	7%	7%	0.40	4.58 %/dB

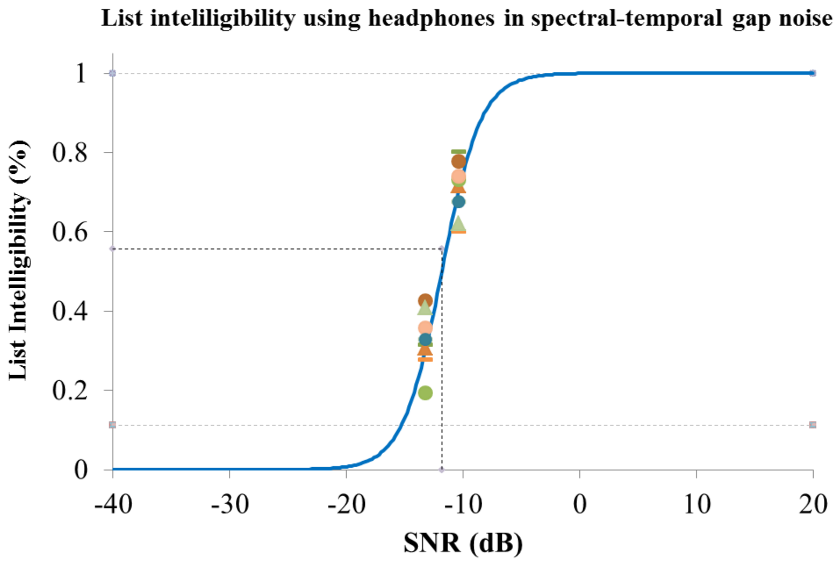


Figure 70: Measured performance-intensity function in noise for all lists in the MDTT using headphones in STG.

The range of SRTn for the MDTT using telephones in TSN was between --10.1 and -10.3 dB SNR, while the slope was between 16.9 %/dB and 23.5 %/dB respectively. The distribution of results measured in fixed SNR for SRTn and slope using telephone in test specific noise are shown in Table 39 and Figure 71 below.

Table 39: Average SRTn and slope for all 8 lists in the MDTT using telephones in TSN.

Telephone in TSN	SNR_1 (%)	SNR_2 (%)	SRT_n	Slope at midpoint:
MalayTripletList1	23%	73%	-10.36	18.41 %/dB
MalayTripletList2	21%	70%	-10.17	18.45 %/dB
MalayTripletList3	23%	69%	-10.20	16.79 %/dB
MalayTripletList4	26%	67%	-10.22	14.60 %/dB
MalayTripletList5	16%	76%	-10.22	23.53 %/dB
MalayTripletList6	23%	70%	-10.23	16.93 %/dB
MalayTripletList7	23%	71%	-10.28	17.49 %/dB
MalayTripletList8	23%	70%	-10.23	17.40 %/dB
Average	22%	71%	-10.24	17.95 %/dB
Standard deviation	3%	3%	0.06	2.56 %/dB

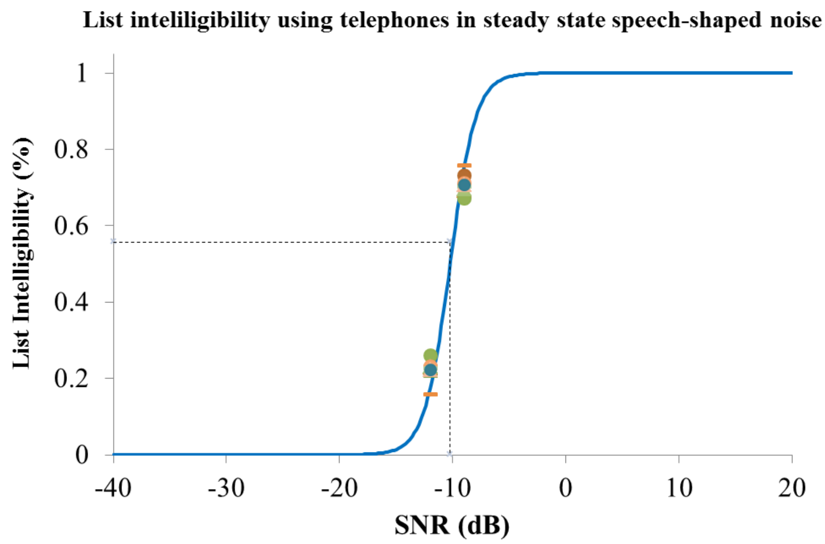


Figure 71: Measured performance-intensity function in noise for all lists in the MDTT using telephones in TSN.

The patterns of differences seen in the headphone application between both background noises can also be observed in telephones. The SRTn using telephones in STG was lower than in TSN, with the slope of intelligibility being slightly shallower in STG than in the TSN. The mean SRTn and slope list-specific intelligibility for all 8 test lists were -10.77 ± 0.25 dB and 17.78 ± 6.7 %/dB, respectively. The distribution of results measured in fixed SNR for SRTn and slope using telephones in the spectrotemporal gap noise are shown in Table 40 and Figure 72 below.

Table 40: Average SRTn and slope for all 8 lists in the MDTT using telephones in STG.

Telephone in STG	SNR_1 (%)	SNR_2 (%)	SRT_n	Slope at midpoint:
MalayTripletList1	33%	61%	-10.84	13.39 %/dB
MalayTripletList2	36%	62%	-11.00	11.92 %/dB
MalayTripletList3	27%	56%	-10.41	13.60 %/dB
MalayTripletList4	33%	58%	-10.72	11.70 %/dB
MalayTripletList5	26%	67%	-10.90	20.19 %/dB
MalayTripletList6	20%	57%	-10.39	19.32 %/dB
MalayTripletList7	22%	72%	-10.95	25.09 %/dB
MalayTripletList8	21%	74%	-10.97	26.99 %/dB
Average	27%	63%	-10.77	17.78 %/dB
Standard deviation	6%	7%	0.25	6.03 %/dB

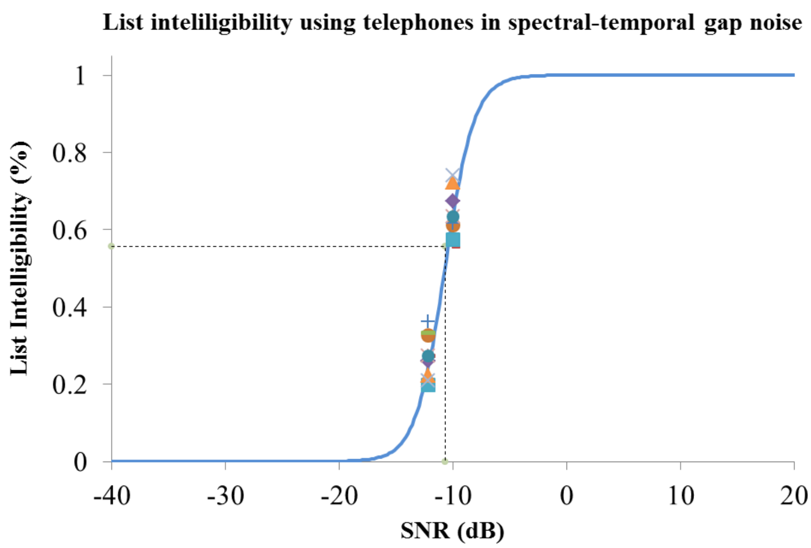


Figure 72: Measured performance-intensity function in noise for all lists in the MDTT using telephones in STG.

Due to the way the pre-set list were arranged, list equivalency for all four test conditions were examined in an omnibus two-way multivariate analysis (two-way ANOVA) to identify any influence of interaction between lists and noise type. As expected, there was a significant influence of using different types of transducers and background noise (where $F(1, 8) = 225.5$, $p < 0.05$ and $F(1, 8) = 66.9$, $p < 0.05$). There were, however, no significant effect of lists and interaction between lists and background noises for all test conditions (where $F(7, 8) = 1.1$, $p = 0.45$ and $F(7, 8) = 1.479$, $p = 0.297$). This suggests that for all four test conditions lists within tests were statistical equivalent.

6.1.3.2 Adaptive measurement

Group 2 participants average hearing thresholds for octave frequencies between 500 Hz to 8000 Hz were 8.55 ± 5.45 dB HL. The figure below shows the hearing threshold distributions for all participants across the audiometric range.

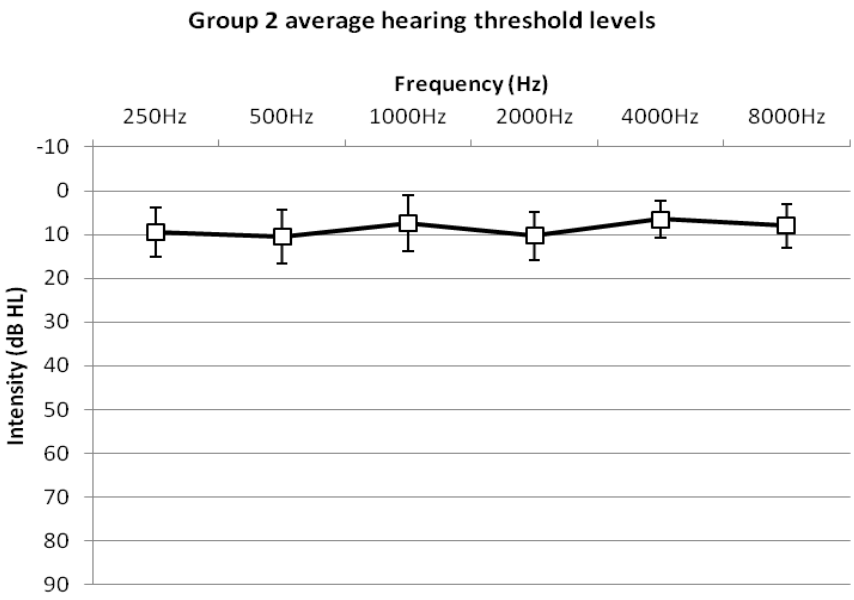


Figure 73: Average hearing threshold measurements for normal hearing participants in Group 2 (n=20).

The table below shows the average SRTn and slope scores obtained using adaptive measures for all four test conditions. Using the adaptive measure, the differences between using the test specific noise and spectrotemporal gap noise using the same type of transducers are less marginal compared to the fixed SNR measurement where differences are less than 1 dB SNR. The SRTn were lower by about 3 dB when headphones were used compared to telephones. The table and figures below show the average SRTn

Table 41: Average SRTn and σSRTn across lists in all four test conditions using adaptive measurement.

		List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	Average
Headphone in TSN	mean	-12.29	-12.56	-12.39	-12.50	-12.33	-12.48	-12.54	-12.46	-12.44
	StDev	1.07	0.72	0.83	0.87	0.89	1.13	1.23	1.02	0.97
	max	-9.80	-11.30	-10.90	-10.40	-10.40	-10.30	-10.60	-10.10	
	min	-14.00	-13.90	-13.70	-13.80	-13.70	-14.00	-14.50	-14.10	
Headphone in STG	mean	-12.61	-12.43	-12.17	-12.71	-12.46	-12.89	-12.99	-13.05	-12.66
	StDev	0.77	0.97	1.10	1.07	0.80	1.00	0.91	0.92	0.94
	max	-11.30	-10.50	-9.60	-10.80	-10.70	-11.10	-11.30	-11.70	
	min	-14.10	-14.90	-14.10	-14.70	-14.00	-14.70	-14.20	-15.00	
Telephone in TSN	mean	-9.12	-9.43	-9.70	-9.47	-9.23	-9.58	-9.26	-9.66	-9.43
	StDev	1.63	1.90	1.41	1.65	1.56	1.80	1.56	1.33	1.61
	max	-5.80	-6.50	-7.40	-6.60	-5.80	-5.10	-7.20	-7.70	
	min	-11.80	-12.60	-11.70	-12.60	-11.60	-12.00	-11.90	-12.60	
Telephone in STG	mean	-8.79	-9.42	-9.60	-9.18	-9.40	-9.21	-9.01	-9.32	-9.24
	StDev	1.95	1.94	1.59	2.15	1.98	2.25	1.84	1.42	1.89
	max	-4.50	-5.70	-6.30	-5.80	-5.50	-3.80	-6.00	-6.10	
	min	-13.10	-12.50	-13.20	-12.50	-13.00	-12.70	-12.00	-11.80	

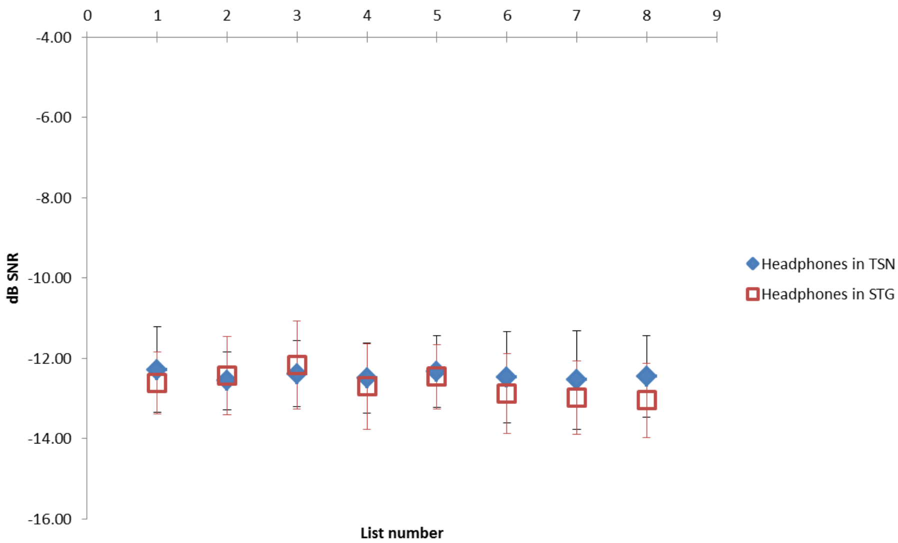


Figure 74: Distribution of SRTn of MDTT lists using headphones and adaptive measurements.

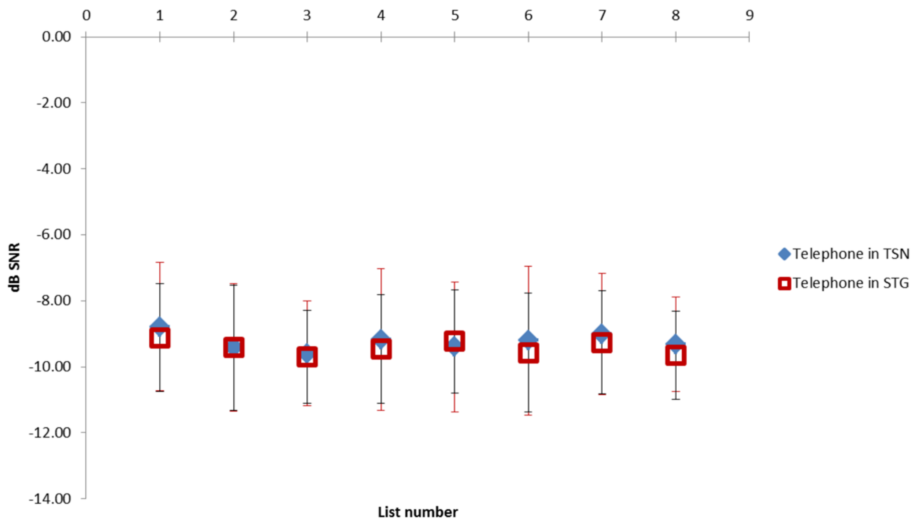


Figure 75: Distribution of SRTn of MDTT lists using telephone and adaptive measurements.

RM-ANOVA was conducted to investigate the influence of lists in the performance of participants. For all four test conditions measured adaptively, no significant effects of lists were found using headphone in TSN ($F_{headphoneTSN}$ (7, 13) = 0.270, p = 0.955); headphones in STG ($F_{headphoneSTG}$ (7, 13) = 1.659, p = 0.204); telephones in TSN ($F_{telephoneTSN}$ (7, 13) = 1.03, p = 0.456); and telephones in STG ($F_{telephoneSTG}$ (7, 13) = 0.687 p = 0.682).

To investigate the effects of noise in all four conditions of tests using adaptive measurements, an additional analysis using RM-ANOVA was conducted by pooling all of the results (n=160) obtained from all participants in Group 2 for all individual test lists (20 participants*8 lists per condition). A significant effect of noise using headphones was observed (where F (1, 159) = 3.929, $p < 0.05$). However no effect of noise was observable when telephones were used (where F (1, 159) = 1.473, p = 0.227).

6.1.4 Discussion

Hearing threshold levels in Groups 1 and 2 were comparable and within normal limits. The total number of participants was limited to 16, but due to the extended testing time for the fixed measurement testing for the MDTT, the testing playlists were split in two, giving an effective sample size of 8. However, the sample size probability (power) was calculated at 0.849 using the Power and sample size calculator software (Dupont & Plummer, 1990) – largely because of the small standard deviation (±0.36 dB SNR using headphone in TSN and ± 0.06 dB SNR using telephone in TSN) between lists.

The measured slopes were lower than predicted, which is consistent with other findings (Jansen et al., 2010; Zokoll et al., 2012). It would seem that there is an overestimation in the prediction of slope scores using Equation 2. Comparisons between transducers show that using telephones reduced the overall sensitivity of the test. This can be directly contributed to the smaller frequency bandwidth of telephone receivers and its poorer audio quality compared to headphones. However, using telephone showed better consistency as the standard deviations of the SRT measurements were reduced when measured at fixed SNRs. This pattern is not clearly seen when the MDTT was measured adaptively.

In the fixed measurement for the MDTT, using headphones in STG noise showed slight improvement (lower SNR) of the average list scores by 0.4 dB SNR compared to TSN. A slightly larger difference was seen between the two noises when telephones were used (0.59 dB SNR). This suggests the potential of using STG noise in improving test sensitivity which will be examined in the next chapter in participants with varying hearing levels. Statistically, using STG noise made a significant difference in these normal hearing listeners in the fixed SNR measurements using both headphones and telephone, and in the headphone condition for the adaptive measure, but not in the adaptively measured telephone condition. These mixed results could be contributed to by the small improvements seen in normal hearing subjects. Although release from masking seen within normal hearing participants were small, this is not necessarily a problem for the test as it is expected that hearing impaired listeners do disproportionately worse than their normal hearing counterparts. As part of the design of the noise, the temporal fluctuations would hopefully reduce the performance of hearing impaired listeners. To test this hypothesis, a series of observations carried out with listeners of varying hearing levels was conducted, and is described in the next chapter.

Compared to other versions of DTT that use TSN as background noise, the mean SRTn for this study was slightly lower compared to the range of other published DTT versions at -11.3 dB SNR. This is also true for the slope of intelligibility, where the slope for the MDTT is shallowest at 15.99 %/dB. The table below shows the comparison between MDTT and several other published versions of DTT.

Table 42: Summary of normative values for normal hearing listeners across various DTT versions. (HP = Headphone, Tel = Telephone)

Test	**Transducer**	**Predicted SRT**	**Predicted Slope**	**SRTn Fixed (across lists)**	**σSRT Fixed (between lists)**	**SRTn Adaptive**	**σSRT Adaptive**	**Slope Fixed (across lists)**	**σSlope Fixed (across lists)**
Malay DTT	HP	-13.4	18.4% /dB	-11.3	0.34	-12.4	0.1	15.9% /dB	2.6% /dB
	Tel	-13.5	17.7% /dB	-10.2	0.1	-9.4	0.2	18.0% /dB	2.6% /dB
French DTT	HP	-9.5		-10.5	0.3	-6.7	0.5	27.1% /dB	3.0% /dB
	Tel	-11.2		-6.4	0.4	-6.4	0.4	17.1% /dB	2.5% /dB
German DTT	HP	-10.2	17.6% /dB	-9.3	0.2			19.6% /dB	
	Tel	-10.5	17.2% /dB	-6.5	0.3			18.0% /dB	
Dutch DTT	HP			-11.2				16.0% /dB	
	Tel			-6.9				20.0% /dB	
Swedish DTT	HP			-6.90				24.20% /dB	
	Tel			-4.30				24.30% /dB	
Polish DTT	HP	-9.40	19.40% /dB	-9.40				19.70% /dB	
	Tel			-7.40					

Statistical analysis suggests that there were no lists differences found in all eight test conditions for both types of measurement. This also suggests that all the lists are usable for further evaluation in listeners with varying levels of hearing. The measurement made using fixed SNR will be applied as the reference norm for normal hearing listeners.

6.2 Evaluation of the Malay auditory-visual matrix sentence test

6.2.1 Introduction

As described in the ICRA recommendations for speech-in-noise tests (Akeroyd et al., 2015), evaluation of the test lists allows assessments of their equivalence and provides normative values for the test. Additionally, due to the slightly more complex nature of the task, significant training effects are seen in the matrix sentence tests. Therefore, an additional study was conducted to identify the magnitude of any training effect seen using the adaptive measurement method in the Malay matrix sentence test. Tests of the equivalency of the matrix sentence test lists were performed separately for both fixed SNR and adaptive measurements. Measurements were only conducted using a closed-set response method where the audio signal was presented monaurally to all participants.

6.2.2 Methods

6.2.2.1 Recruitment of participants

Three groups of participants were recruited to evaluate the matrix sentence test in fixed SNR measurements, adaptive measurements and in a study to investigate training effects in the MMST-AV.

Group 1 participants were tasked to perform the test using fixed SNR measurements. They consisted of 10 adult Malay native speakers (3 male and 7 female) aged 22.5 ± 2.3 years old who were students at the International Islamic University Malaysia. Participants were paid RM50 for their time and effort. Testing was broken into two sessions and the overall testing time ranged from five to seven hours per participant. All Group 1 participants had normal hearing with hearing threshold levels below 20 dB HL at all octave frequencies.

Group 2 participants were the same group that performed the digit evaluation in the adaptive approach described in Section 0 above. They consisted of 20 adult Malay native speakers aged 26.4 ± 6.9 years old (5 male and 15 female). They were paid an additional RM50 for their time and effort.

Group 3 participants were assigned to perform the training effect investigation study. 10 adult Malay native speakers aged 21.6 ± 1.6 years old were recruited (4 male and 6 female undergraduate students of the International Islamic University Malaysia). All Group 3 participants had hearing thresholds below 20 dB HL at all octave frequencies. They were also paid RM 50 for their time and effort.

6.2.2.2 Test procedure I: Training effects

Training effects were studied in a separate group of normal hearing listeners. Participants in this group had no prior experience in this test. Participants were asked to complete five test lists sequentially, both in TSN and BN. Half of the participants would perform the test in TSN first and then BN, and vice versa for the other half. The subsequent measurement of the other half was done a week after the first measurement. The test was conducted adaptively using the dual track adaptive measurement procedure as mentioned above and explained in detail in Section 6.2.2.4. Participants were briefed and were requested to click using a mouse on the word selection provided in the software interface. They were encouraged to guess if they were unsure.

6.2.2.3 Test procedure II: Fixed SNR measurement

Testing was carried out in a double walled audiometric cabin at the IIUM Hearing & Speech clinic at the International Islamic University Malaysia, Kuantan Campus. Participants were seated in front of the computer and were asked to click on the words they heard during the test. A display listing word choices were presented as soon as the each sentence is complete. Participants were allowed as many breaks as they please using a designated pause button. Test was carried monaurally at a constant 65 dB background noise and the desired SNR was achieved by adjusting the level of stimuli. 15 test lists of 30 sentences were used for evaluation purposes. All participants were given 2 training list before the any measurement is taken. For the steady state speech-shaped noise or test specific noise (TSN) the lists were measured at -12.7 and -6.3 dB SNRs. The test in 6-talker babble noise was conducted in -8.4 and -2.1 dB SNRs. The fixed SNR levels were predetermined using the normalisation data and a pilot study in 3 normal hearing subjects (including author). SRTn and slope scores were identified using Equation 3.

6.2.2.4 Test procedure II: Adaptive measurement

Testing was carried out that the same location as in the fixed SNR measurements. Participants were required to respond in the same way as in the fixed measurement procedure. To arrive at the SRTn and record the slope of intelligibility, a dual track adaptive procedure using varying step sizes and word scoring was adapted from that described in Brand & Kollmeier (2002), and modified according to recommendations by Brand (T. Brand, personal communication, February 3, 2015). The modified procedure is as follows:

where , = reversal number, and . The target intelligibility was set to 80% for track 1 and 20% for the track 2. The "speed factor" was not permitted to fall below a minimum value of 0.25. In addition, was doubled if i) both and were 80% or greater for track 1 or 20% or lower for track 2; and ii) was greater than 0.5.

A programming error in which was set to "number of trials" instead of "number of reversals" caused a slower rate of convergence on the targets – however, the number of trials presented ensured that the data gathered around each target enabled adequate fitting of the psychometric functions, which was done using a Levenberg-Marquardt nonlinear regression. This fit to the data gathered around the targets at 20% and 80% provided the speech reception threshold (which was set at the 50% score of intelligibility) and the slope of the function in %/dB or $^{-1}$dB using Equation 3.

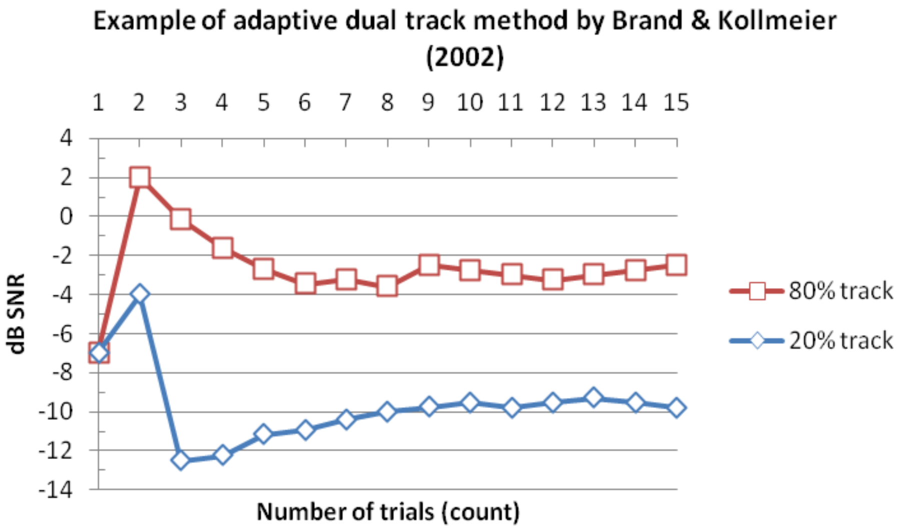


Figure 76: Illustration of the dual track adaptive method used in the MMST-AV as proposed by Brand & Kollmeier (2002). Equal number of sentence trials were done for each track. Step size is decreased as the number of trials proceeds converging to the estimated target correct response of 20% or 80%. The convergence on targets was slower than intended, but nonetheless, both tracks were within 1 dB of their targets by Trial 9 of 15.

6.2.3 Results and analyses

6.2.3.1 Training effects

Figure 77 below shows mean SRTn as a function of temporal order for the MMST-AV in TSN. RM-ANOVA indicated that there was a significant effect of test order or training effect in TSN ($F(4, 6) = 11.64$, $p < 0.05$). As in other MST versions, greatest difference was found between the first presented test list (referred to as List 1 here) and the second (List 2) at –1.45 dB. The difference between List 2 and List 5 was 0.6 dB. Multiple pairwise comparisons with Bonferroni correction showed that the measurements of List 1 differed significantly ($p < 0.05$) from the other lists. The average SRTn from List 2 to List 5 was recorded at -10.4 ± 0.6 dB SNR.

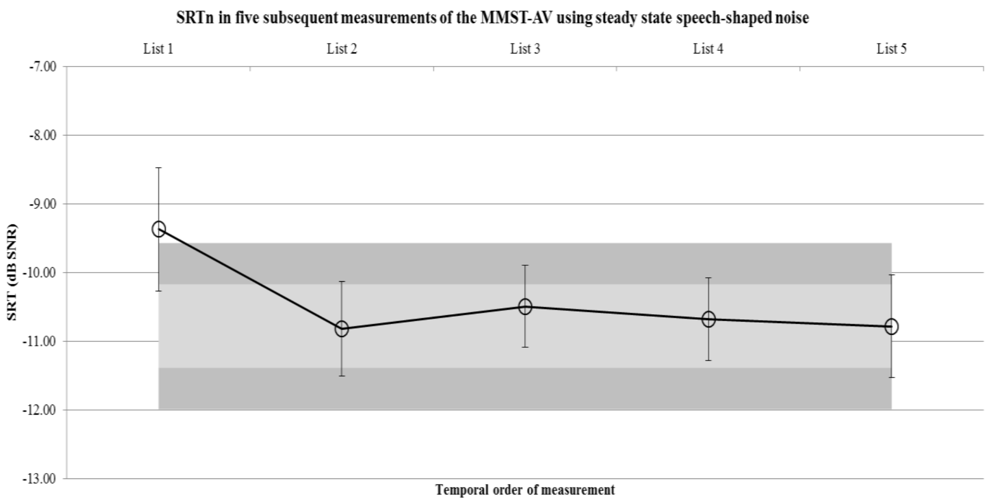


Figure 77: Mean SRTn using TSN with corresponding standard deviation as a function of measurement number. Dark grey shade indicates values within 2 standard deviations of average SRTn across lists and light grey shade indicates values within 1 standard deviation.

A similar training effect pattern was observed for the MMST-AV in BN. To investigate this statistically, an RM-ANOVA was conducted and revealed a significant effect of training effect (where $F(4, 6) = 7.53$, $p < 0.05$). Pairwise comparison was conducted with Bonferroni correction which showed a significant difference between measurements in the first list and subsequent list measurements. The first list differed from the second by 0.8 dB whereas the difference between the second list and the last list was 0.5 dB. The average SRTn from List 2 to List 5 for the the in BN was recorded at -6.5 ± 0.6 dB SNR.

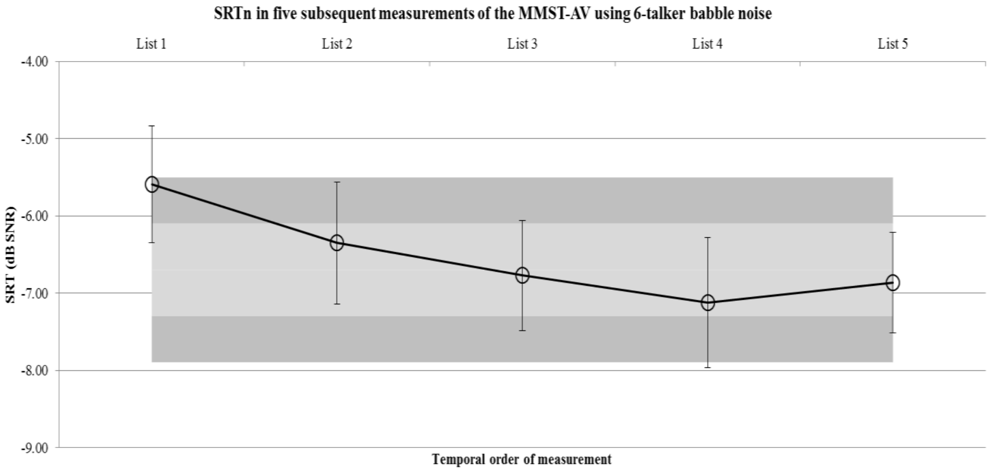


Figure 78: Mean SRTn using BN with corresponding standard deviation as a function of measurement number. Dark grey shade indicates values within 2 standard deviations of average SRTn across lists and light grey shade indicates values within 1 standard deviation.

6.2.3.2 Fixed SNR measurement

No participants reported any history of hearing or ear problems. The average hearing threshold of all participants (n = 10) was 10.4 ± 4.3 dB HL for all octave frequencies. The distribution of hearing threshold levels for participants in Group 1 is as shown in Figure 79 below.

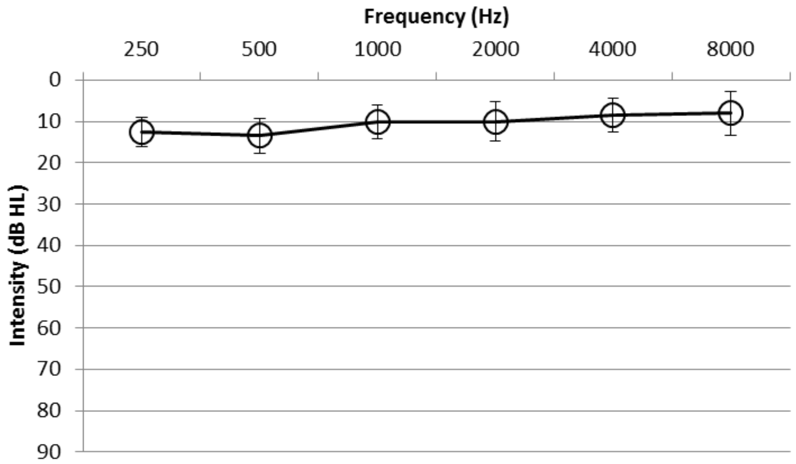


Figure 79: Group 1 average hearing threshold levels (n=10).

For the fixed SNR test in TSN, the mean SRTn and slope averaged across all 15 lists were -10.1 ± 0.2 dB SNR and 14.9 ± 1.2%/dB, respectively. The lowest recorded SRTn across lists was -10.3 dB (List 15) and the highest was -9.6 dB (List 6); the lowest slope across lists was 13.3%/dB (List 9) and the highest was 17.2%/dB (list 4). To determine the influence of lists in the MMST-AV using TSN, an RM-ANOVA was conducted on all data. No statistical difference was detectable for the effect of lists for SRTn (F (4.835, 43.51) = 1.165, p = 0.252, Greenhouse-Geisser correction) or the slope (F (4.847, 43.63) = 1.04, p = 0.404). Shown below Table 43 are the SRTn and slopes for the lists in the MMST-AV using test specific noise.

Table 43: Average SRTn and slope for all 15 lists in the MMST-AV in TSN measured in fixed SNRs.

TSN	SNR_1 (%)	SNR_2 (%)	SRTn	σSRT	Slope (%/dB)	σSlope (%/dB)
List 1	20%	88%	-10.1	1.3	15.2	3.2
List 2	18%	89%	-10.0	0.8	16.7	2.0
List 3	22%	88%	-10.2	0.9	14.6	4.2
List 4	17%	89%	-9.9	1.2	17.2	2.1
List 5	19%	85%	-9.8	0.8	15.1	4.3
List 6	17%	85%	-9.6	0.8	15.9	2.5
List 7	21%	90%	-10.3	0.9	15.8	2.4
List 8	21%	89%	-10.2	0.9	15.3	2.5
List 9	24%	86%	-10.2	1.1	13.3	2.0
List 10	21%	90%	-10.2	1.1	15.7	3.2
List 11	23%	85%	-10.1	1.0	13.4	2.2
List 12	21%	87%	-10.1	1.4	14.3	3.0
List 13	25%	88%	-10.5	1.0	13.7	2.0
List 14	22%	86%	-10.1	0.9	13.8	2.5
List 15	23%	88%	-10.3	0.9	14.0	2.0

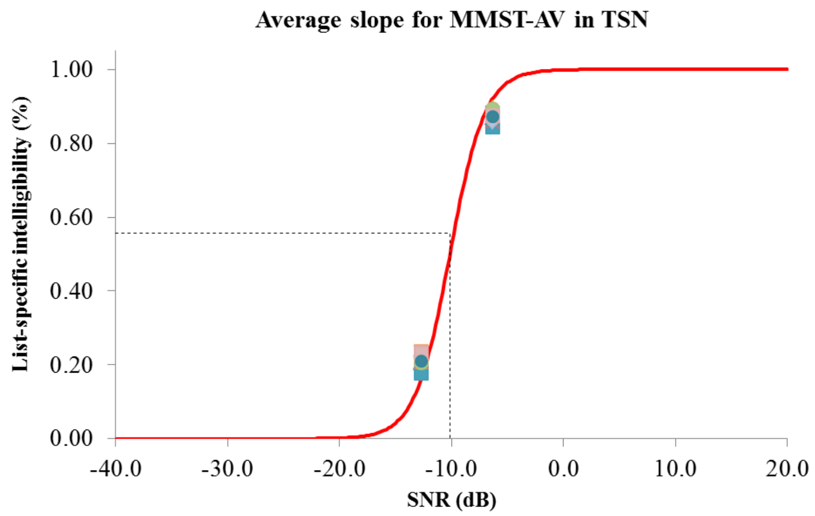


Figure 80: Average slope across all 15 lists in the MMST-AV in TSN

As expected the test in BN showed higher SRTn and shallower slope as compared to the TSN. The mean SRTn and slope averaged across all 15 lists for MMST-AV in BN were –6.4 ± 0.2 dB SNR and 12.2 ± 0.7%/dB, respectively. The lowest recorded SRTN across lists was -6.0 dB (List 1 & 3) and the highest was -6.7 dB (Lists 6 &14); the lowest slope across lists was 11.4%/dB (List 2 & 11) and the highest was 13.3%/dB (List 5). To statistically test list equivalency, an RM-ANOVA was conducted on all data. No statistical difference was observable for the effect of lists for SRTn ($F(2.857, 25.7) = 8.46$, $p = 0.052$, Greenhouse-Geisser correction) or slope ($F(3.04, 27.37) = 3.811$, $p = 0.21$). Shown below are the SRTn and slope for the MMST-AV using 6-talker babble noise.

Table 44: Average SRTn and slope for all 15 lists in the MMST-AV in BN measured in fixed SNRs.

Babble Fixed SNR	SNR_1 (%)	SNR_2 (%)	SRTn	σSRT	Slope (%/dB)	σSlope (%/dB)
List 1	19%	83%	-6.0	0.4	12.0	2.8
List 2	23%	84%	-6.3	1.1	11.4	1.3
List 3	17%	85%	-6.0	0.4	13.2	2.1
List 4	20%	87%	-6.4	0.9	13.0	3.7
List 5	21%	88%	-6.5	1.1	13.3	5.7
List 6	24%	87%	-6.7	1.0	12.2	3.2
List 7	22%	87%	-6.5	0.3	12.5	1.2
List 8	24%	85%	-6.5	1.4	11.5	3.6
List 9	22%	84%	-6.3	0.7	11.6	2.9
List 10	25%	86%	-6.6	0.8	11.7	4.9
List 11	23%	84%	-6.3	0.8	11.4	3.2
List 12	23%	85%	-6.4	0.8	11.7	3.2
List 13	23%	87%	-6.6	1.0	12.5	2.0
List 14	23%	89%	-6.7	1.1	12.9	2.3
List 15	23%	87%	-6.5	0.9	12.3	2.1

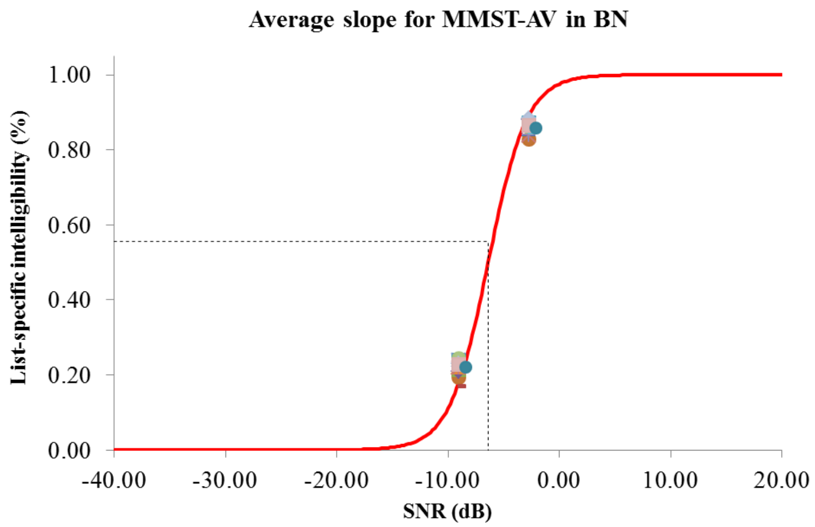


Figure 81: Average slope across all 15 lists in the MMST-AV in BN.

The average SRTn for the MMST-AV in TSN was 3.69 dB lower than in BN. The slope in BN was shallower by 2.7%/dB compared to the slope in TSN. The figure below illustrates the differences between the two measurements for SRTn and slope.

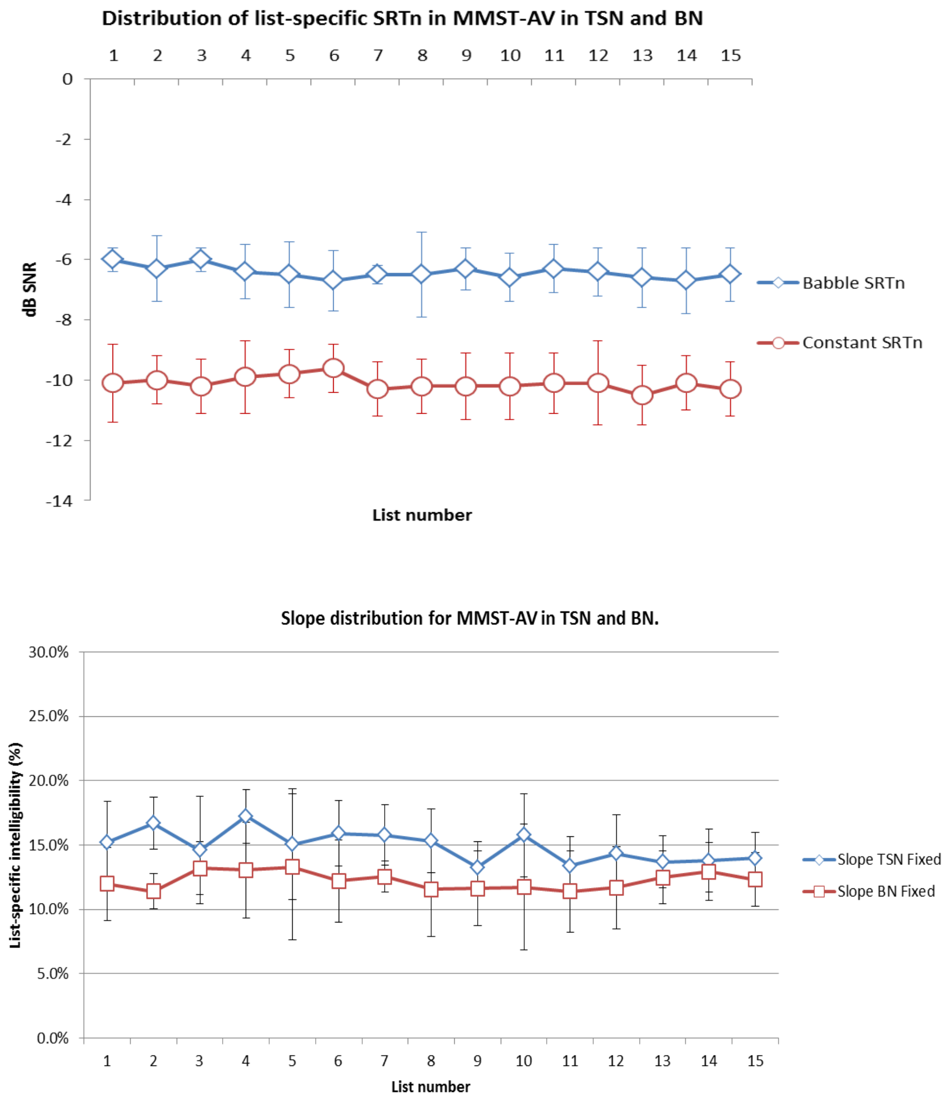


Figure 82: Comparison between SRTn (top) and slope (bottom) for MMST-AV in test specific noise and 6-talker babble noise measured in fixed SNRs.

6.2.3.3 Adaptive measurements

For the description of Group 2 audiometric results, please refer to the Results section for the MDTT adaptive evaluation (Section 0 above).

Adaptive measurement for the MMST-AV in TSN showed average SRTn and slope estimates across all 15 lists of -10.77 ± 0.4 dB and 11.53 ± 0.37%/dB, respectively. The lowest recorded SRTn was -11.4 ± 2.4 dB (List 12) and the highest was -10.1 ± 1.4 dB (List 3); the shallowest slope for the MMST-AV in TSN was 9.7 ± 2.3%/dB (List 11) and the steepest was 13.2 ± 3.7%/dB (List 7). To statistically investigate the equivalency of the lists, an RM-ANOVA was conducted and revealed no significant list effect for either SRTn (F (4.4, 84.93) = 2.395, p = 0.05, Greenhouse-Geisser correction). To compare the effects of slope between lists, Friedman non parametric repeated measure ANOVA showed no significant differences were observable where, $\chi^2(14) = 13.72$, p = 0.471. Test-retest reliability –defined as the root mean square of the within-listeners standard deviation for repeatedly measured adaptive SRTn (Jansen et al., 2012) – was measured at 1.1 dB, meaning that in 95% of cases then measured SRTn differs from the subjects true SRTn by less than 2.2 dB (Plomp & Mimpen (1979)). Shown below in Table 45 is the distribution of SRTn and slope for the MMST-AV in TSN.

Table 45: Average SRTn and slope for all 15 lists in the MMST-AV in test specific noise measured adaptively.

TSN	SRTn	σSRT	Slope (%/dB)	σSlope (%/dB)
List 1	-10.2	1.6	12.0	3.6
List 2	-10.3	1.9	12.5	3.3
List 3	-10.1	1.4	11.8	2.3
List 4	-10.9	2.2	11.2	3.6
List 5	-10.2	1.9	11.8	3.8
List 6	-11.2	2.5	12.0	4.5
List 7	-10.6	1.7	13.2	3.7
List 8	-10.8	1.3	10.8	2.1
List 9	-10.9	2.8	11.0	2.7
List 10	-11.1	1.8	12.0	2.9
List 11	-11.0	2.4	9.7	2.3
List 12	-11.4	2.4	11.4	3.7
List 13	-11.2	2.3	11.9	3.9
List 14	-11.1	1.9	11.0	2.6
List 15	-10.6	2.0	10.8	2.7

The mean SRTn and slope averaged across all lists MMST-AV using BN were -7.7 ± 0.37 dB and 11.95 ± 1.18%/dB, respectively. The lowest recorded SRTn was -9.1 ± 8 dB (List 8) and the highest was -7.5 ± 6.8 (List 1); the shallowest slope for the MMST-AV in BN was 9.7 ± 2.3%/dB (List 11) and the steepest was 13.2 ± 3.7%/dB (List 7). To statistically investigate the lists equivalency, an RM-ANOVA was conducted and revealed no significant list effect that was found in the test within-subjects effects for SRTn ($F(6.83, 129.66) = 2.426$, $p = 0.054$, Greenhouse-Geisser correction). For the effect lists on slope scores, a significant effect was detectable using Friedman non parametric repeated measure ANOVA where, $\chi^2(14) = 61.73$, $p < 0.05$. Pairwise comparison using Wilcoxon signed-rank test showed significant differences between lists 1 and 13 ($Z = -2.59$, $p < 0.05$) and lists 1 and 8 ($Z = -2.803$, $p < 0.05$).

Table 46: Average SRTn and slope for all 15 lists in the MMST-AV in 6-talker babble noise measured adaptively.

BN	SRTn	σSRT	Slope (%/dB)	σSlope (%/dB)
List 1	-7.5	6.8	11.7	2.9
List 2	-8.2	7.5	11.1	2.8
List 3	-8.2	7.5	11.8	3.0
List 4	-8.7	8.0	10.9	2.5
List 5	-8.4	7.7	11.3	3.1
List 6	-8.2	7.5	14.2	4.4
List 7	-8.7	8.0	11.6	3.1
List 8	-9.1	8.4	9.9	3.5
List 9	-8.3	7.6	12.2	3.2
List 10	-8.6	7.9	13.8	5.6
List 11	-8.3	7.6	13.3	4.4
List 12	-8.0	7.3	12.3	2.9
List 13	-8.3	7.6	11.8	2.6
List 14	-8.2	7.5	12.6	3.2
List 15	-7.9	7.2	10.7	2.9

The SRTn difference between TSN and BN when measured adaptively is 3.1 dB (lower in BN) – a value very similar to the difference observed using fixed SNR (3.7 dB lower in BN). The slope in BN was very similar to that of TSN, with a reduction of only 0.4%/dB compared to the slope in TSN. Figure 83 belowillustrates the differences between the two measurements for SRTn and slope.

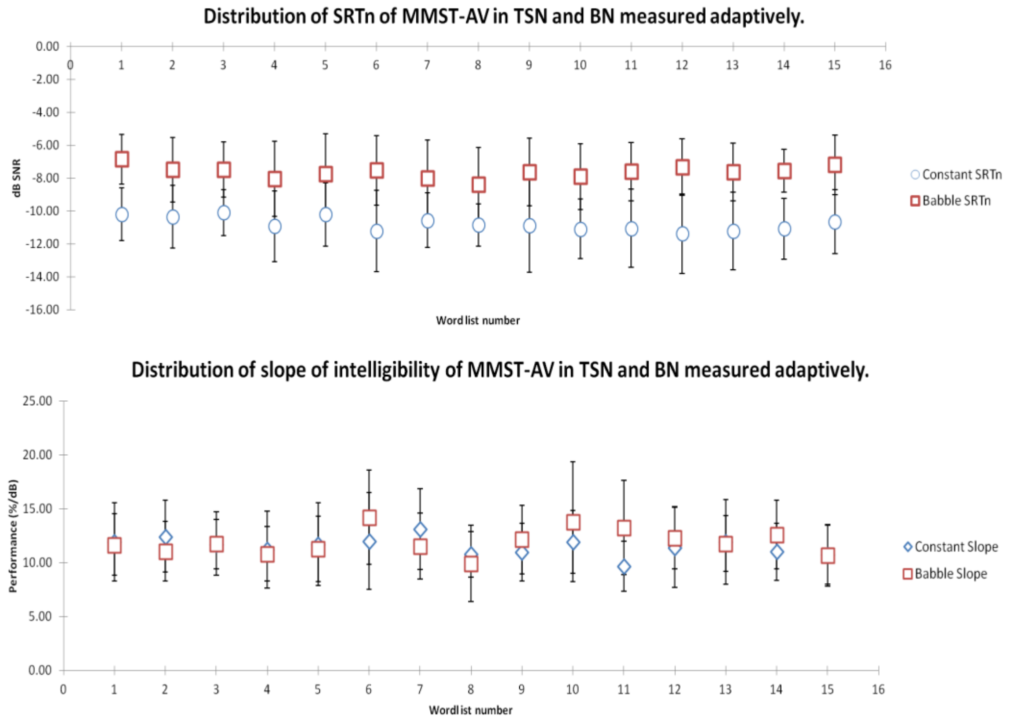


Figure 83: Comparison between SRTn (top) and slope (bottom) for MMST-AV in test specific noise and 6-talker babble noise measured adaptively.

6.2.4 Discussion

The evaluation of the MMST-AV in fixed SNR verified the improvement in overall test intelligibility, as the steepness increased by 1.5% from pre-normalisation slope of 13.4%/dB to measured post evaluation slope of 14.9%/dB. The improvement resulting from the normalisation process is marginal compared to those for the Danish MST (Wagener et al., 2003), Italian MST (Puglisi et al., 2015), Polish MST (Ozimek et al., 2010) and Spanish MST (Hochmuth et al., 2012) which gained 3.9, 3.7, 3.2 and 2.2%/dB of measured improvement after normalisation, respectively. However, the improvement of slope in the MMST-AV is similar to the reported benefit described in the German MST which was 1.2%/dB (Wagener, Brand, et al., 1999). Puglisi et al. (2015) suspects that the amount of improvement could be proportionate to the maximum level correction applied for each test. For example, the German MST was adjusted by maximum of ± 2 dB whereas the Danish MST was adjusted by up to ± 4 dB. There could be some grounds for this assumption, however other factors such as speaker influence cannot be ruled out for now. As for the prediction slope value, a pattern similar to that observed in the MDTT was also seen here, as an overestimation of slope scores was found in the MMST-AV predictions: The predicted slopes for the MMST-AV in TSN and BN were 18.5 and 11.1%/dB whereas the measured slopes were 14.9 and 12.2%/dB. However, the difference in SRTn between lists seen in the fixed SNR measurement is comparable to other MST tests at standard deviation of 0.2 dB across lists (Hochmuth et al., 2012; Ozimek et al., 2010; Puglisi et al., 2015; Wagener, Brand, et al., 1999; Wagener et al., 2003; Warzybok, Zokoll, et al., 2015).

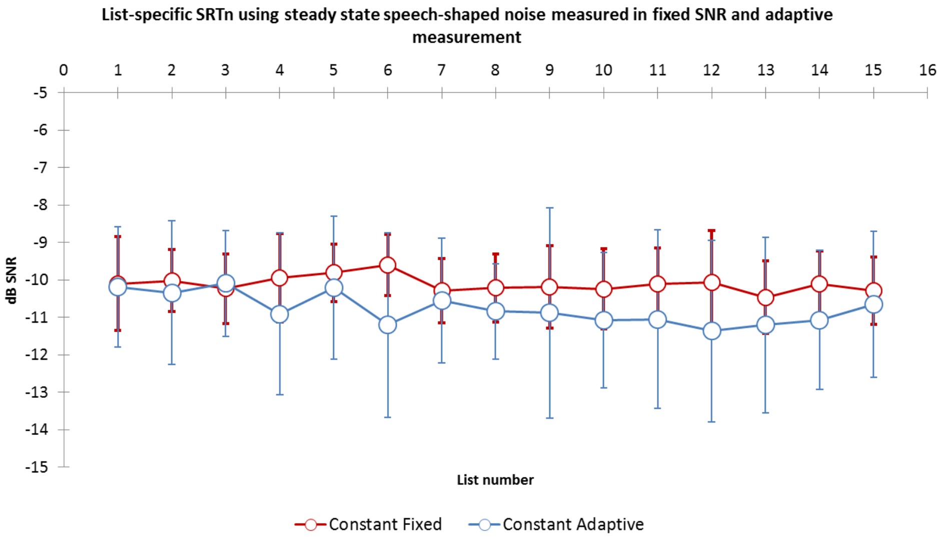
In terms of slope of intelligibility, the MMST-AV showed a mean slope across lists of 14.9 ± 1.2%/dB, which is between the reported slope scores of those observed for MSTs in other languages. The highest slope of intelligibility was reported for the Polish MST (Ozimek et al., 2010) at 17.1%/dB and the lowest was for the German MST (Wagener, Brand, et al., 1999) at 12.2%/dB. As for the mean SRTn averaged across lists, the MMST-AV shared the lowest SRTn with the Finnish MST (as reported in Kollmeier et al. (2015) at -10.1 dB SNR. The highest SRTn was reported in the French version (Jansen et al., 2012) at -6.0 dB SNR.

Comparison of SRTn and slope scores of those observed in MMST-AV and for other MST in other languages is shown in the table below.

Table 47: Comparison and summary of SRTn and slope of intelligibility for MMST-AV and MMST in other languages using closed-set response.All the MST below were spoken by a female speaker except for the Polish MST.

	SRTn Fixed (across lists) dB	σSRT Fixed (between lists) dB	SRTn Fixed (across listeners) dB	σSRT Fixed (between listeners) dB	SRTn Adaptive dB	σSRT Adaptive dB	Slope Fixed (across lists) % /dB	σSlope Fixed (across lists) % /dB
Malay	-10.1	0.2	-10.1	0.99	-10.4	0.6	14.9	1.2
NZ English	-10.7	0.2					10.6	
German	-9.3	0.16			-6.8		12.5	1.2
Danish	-8.4	0.16					12.6	0.9
Polish	-9.6	0.2			-8	1.3	17.1	1.6
Turkish	-8.2	0.2	-8.3	0.8	-7.2	0.8	14.7	1.7
French	-6	0.1	-6	1.1			14.0	
Spanish	-7.7				-7.2	0.7	14.0	
Russian	-9.5	0.2	-9.4	0.7	-9.4	0.8	13.8	1.6
Italian	-7.3	0.2	-7.4	0.9	-7.4	0.8	13.3	1.2
Finnish	-10.1	0.1	-10.1	0.7	-9.7	0.7	16.7	1.2
Average	-8.7	0.2	-8.6	0.9	-8.3	0.8	14.4	1.3

The fixed and adaptive measurements of both SRTn and slope for the MMST-AV differ by a small margin. The fixed SRTn was -10.1 dB whereas the adaptive was -10.4 dB. This pattern is similar from the rest of the other MSTs as they reported the fixed SNR measurement as either being lower or the same with the SRTn in the adaptive measurement. The range of difference between fixed SNR measurements and adaptive measurements in other MST versions was between 0 to 0.9 dB. The adaptive measurement average was taken by removing the first list from the training effect investigation and then averaging subsequent measurements. By applying the default method used in other MST, the SRTn in adaptive measurement was found to be -10.4 ± 0.61 dB SNR which is still very similar to the investigation done in the full set of adaptive measurements in all of the 15 lists. The difference of 0.6 dB between the fixed and adaptive measurement in the MMST-AV is still within the 0.9 dB difference shown in other MSTs. Figure 84 below illustrates the differences between SRTn and slope in the MMST-AV across all lists using fixed SNR and adaptive measurements.



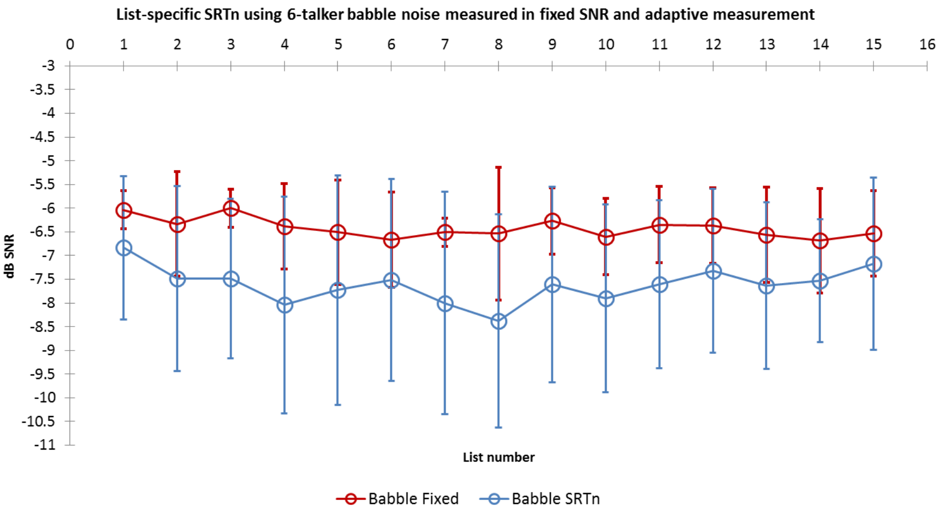


Figure 84: Comparison between SRTn obtained using fixed measurement for test specific noise (top) and 6-talker babble noise (bottom).

Training effect was observed in both types of background noises for the MMST-AV which is consistent with other versions of the MST using TSN. The biggest difference between measurements was recorded between the first and second presented lists. As with other MSTs, two practice lists are recommended before any measurements are taken for the MMST-AV. The small standard deviation observed in both fixed and adaptive measurements show that the test lists are equivalent and homogenous. It is important here to note that even with unequal distribution of words between lists, the list-specific SRTn values across all test lists did not show any statistical difference. This would suggest that the random organization of words within lists did not affect the overall outcome of list-specific intelligibility because of the normalisation process (where specific level adjustments were made to the speech stimuli) and the systematic arrangement of words and sentences that gave equal estimated slope score per list. It is possible to revisit the arrangement of words to make it more equal, however level adjustment is suspected to play a more significant role in allowing test list to be equally difficult, as discussed in Smits & Houtgast (2005).

The MMST-AV showed comparable results to the Malay HINT (Quar et al., 2008) where the monaural unilateral presentation SRTn in the HINT was -12.4 ± 1.0 dB compared to a higher SRTn for the MMST-AV.at -10.1 ± 0.2 dB. Steepness of the performance-intensity slope for the MMST-AV is higher at 14.9%/dB compared to 9.8%/dB for the Malay HINT. Comparisons between languages between the HINT and MSTs showed that as a general observation most MST produced steeper slopes at an average of 14.4%/dB (see Table 47) compared to 10.3%/dB (Soli & Wong, 2008). As the generated test lists were found to be homogenous in both fixed SNR and adaptive measurements, they will be used for the purpose of validation of the test in listeners with varying levels of hearing.

CHAPTER 7

VALIDATION OF THE DIGIT TRIPLET AND AUDITORY-VISUAL MATRIX SENTENCE TEST IN MALAY

7.1 Investigation of the Malay digit triplet test in normal and hearing impaired listeners

7.1.1 Introduction

The digit triplet test has been shown to produce good levels of sensitivity and specificity in screening for hearing loss by measuring speech perception in noise abilities and comparing it to test-specific cut-off limits to determine if the hearing status is ‘good’, ‘insufficient’ or ‘poor’. Smits & Houtgast (2005) reported that the Dutch DTT used limits of <-4.1, -4.1≤ SRTn ≤-1.4 and >-1.4 for good, insufficient and poor respectively. Participants who were found to have a hearing status of ‘good’ were recommended to only seek professional help when in doubt. Participants with ‘insufficient’ were advised to seek help, whereas those with ‘poor’ hearing status were strongly advised to seek help for their hearing (Smits & Houtgast, 2005). In a later study, participants who were found to have ‘poor’ hearing status were found to be 10% more likely to seek help for their hearing compared to those determined to have ‘insufficient’ hearing status (Smits, Merkus, et al., 2006). The cut-off limits for the French DTT are very close to those of the Dutch DTT. In the French DTT, for ‘good’ hearing status the limit is <-4 dB SNR whereas the borderbetween ‘insufficient’ and ‘poor’ was set at -2.7 dB SNR (Jansen et al., 2010). For the American version, a SRTn cut-off of <-5.7 dB was used for the normal hearing and SRTn beyond this value as the refer criteria (Watson et al., 2013). This produced a test sensitivity and specificity of 80% and 83%, respectively. Based on the data above, the cut-off limits between DTT versions in other languages are quite similar, and produce similar levels of test sensitivity and specificity. In this study, the MDTT was validated by investigating responses obtained from listeners with varying levels of hearing. Using this data, the optimum cut-off levels for the MDTT were determined using the receiver-operator curve calculation (ROC) method.

7.1.2 Methods

7.1.2.1 Recruitment of participants

46 participants were recruited for the validation of the MDTT and the MMST-AV in auditory-alone mode. 20 participants had normal hearing levels (average HTL at 500, 1000, 2000 & 4000Hz of < 20 dB HL) and 26 participants had varying degrees of sensorineural hearing loss (average HTL at 500, 1000, 2000 & 4000 Hz of 42.4 ± 15.73 dB HL). Participants were recruited using convenience sampling. Participants were staff, students or patients at the IIUM Hearing & Speech clinic, International Islamic University Malaysia, Kuantan campus. Participants were paid RM50 for their time and effort.

7.1.2.2 Test procedure

This stage of study was conducted at the IIUM Hearing & Speech Clinic at the International Islamic University Malaysia, Kuantan campus. Hearing thresholds were recorded at all octave frequencies before the measurements starts. Participants were seated in a double walled audiometric cabin and were asked to respond by using the keyboard to choose the digit triplets heard. The SRTn was measured adaptively using 2 dB step sizes and responses were used to compare speech perception performance between groups. Digit triplets were presented monaurally to with background noise fixed at 65 dB and stimuli varied across test procedure to achieve target SNR. Each participant were given one random list to complete each for all four test conditions (Headphone using TSN, headphone using STG, telephone using TSN and telephone using STG). The digit triplet test was presented using the University of Canterbury adaptive speech test platform (UCAST) on a Windows™ PC using an external sound card (Creative X-Fi 5.1 SoundBlaster sound card). For the headphone and telephone applications, the transducers Sennheiser HD 280 Pro headphone and Cisco unified series 7900 telephone were used. The telephone handset was coupled to the sound card via a JK Audio THAT-2 audio handset tap.

ROC curves were generated to determine sensitivity and specificity, and optimum cut-off limits for ‘good’, ‘insufficient’ and ‘poor’ hearing categories. For ‘good’ hearing, the classification was predetermined at the upper 95% confidence interval for normal hearing (average hearing threshold at 500, 100, 2000 & 4000 Hz of < 20 dB HL). The second border (which is between 'insufficient' and 'poor' speech perception status) is set by adding up two standard deviations.

7.1.3 Results and analysis

As shown in the figure below, the distribution of hearing levels within the hearing impaired group adequately covers a large range in the audiogram. One participant from the normal hearing groups had a hearing threshold of 40 dB HL at 8000Hz. Since that was the only frequency of theirs to fall outside of the normal hearing criteria, the participant was still in included in the normal hearing group. Both groups were not aged matched as there was a significant difference in age ($t(19) = 9.1$, $p < 0.00$) between the hearing impaired group (mean age of 50.31 ± 11.62 years) and normal hearing group (mean age of 26.4 ± 6.91 years). The age distribution is shown in detail in Table 48 below. Therefore, age will be taken into account in all inferential analysis as it is also known that performance of speech in noise deteriorates with age in adaptive test procedure (Dubno, Dirks, & Morgan, 1984; Mukari, Haniza, Wahat, & Mazlan, 2014).

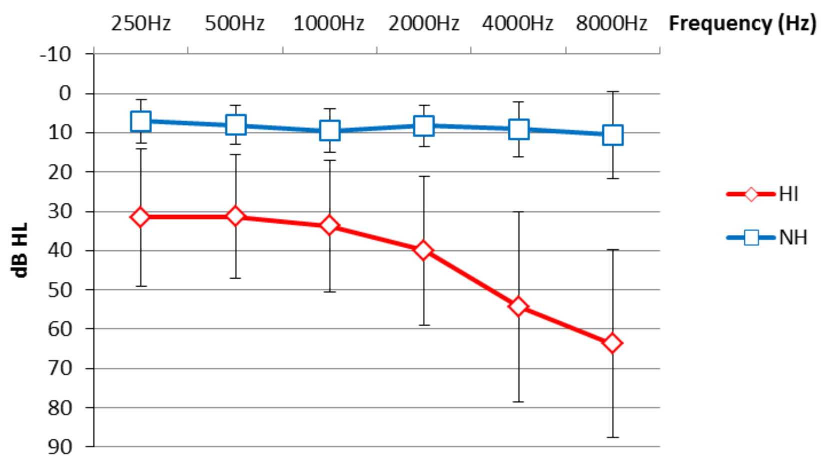


Figure 85: PTA hearing threshold average at octave frequencies for normal (blue squares) and impaired hearing subjects (red diamonds)

Table 48: Distribution of participants average hearing thresholds, age and gender.

Hearing threshold averages across octave frequencies from 250 to 8000 Hz	Age group	Gender	
		Male	Female
0 to 20 dB HL	18 to 29	4	14
	30 to 39	0	1
	50 to 59	1	0
25 to 45 dB HL	30 to 39	5	0
	40 to 49	1	1
	50 to 59	4	1
	60 to 69	6	0
50 to 65 dB HL	18 to 29	0	1
	40 to 49	1	0
	50 to 59	1	2
70 to 85 dB HL	40 to 49	1	0
	50 to 59	2	0
	Total	26	20

As expected, the hearing impaired group showed higher averaged SRTn across all four test conditions. Additionally, responses obtained using telephones were higher in both the normal and impaired hearing groups. The mean score averaged across normal hearing participants using headphone in STG noise at -14.7 dB SNR was the lowest across all four test conditions whereas the highest average score between groups was for the hearing impaired tested using telephone in STG noise at 1 dB SNR.

Table 49: Table showing average, maximum and minimum SRTn values between the normal and hearing impaired groups in all four test conditions.

Noise		Test specific noise		Spectrotemporal gap noise	
Transducer		Headphone SRTn (dB SNR)	Telephone SRTn (dB SNR)	Headphone SRTn (dB SNR)	Telephone SRTn (dB SNR)
	Mean	-6.6 ± 0.4	-0.6 ± 0.6	-5.7 ± 0.7	1.0 ± 0.6
HI	Max	3.0 ± 0.9	12.0 ± 1.1	5.2 ± 3	12.0 ± 1.1
	Min	-11.7 ± 0.1	-7.9 ± 0.2	-12.3 ± 0.2	-9.3 ± 0.3
	Mean	-12.2 ± 0.4	-9.5 ± 0.5	-12.4 ± 0.4	-9.3 ± 0.4
NH	Max	-10.2 ± 1	-4.6 ± 0.9	-9.6 ± 0.6	-4.1 ± 0.8
	Min	-13.9 ± 0.1	-13 ± 0.2	-14.7 ± 0.1	-12.6 ± 0.2

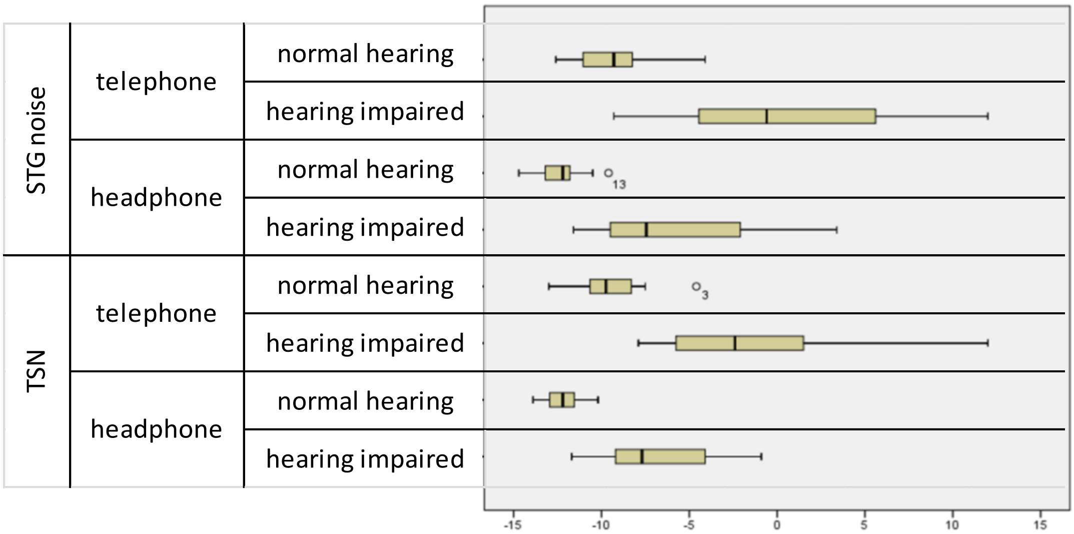
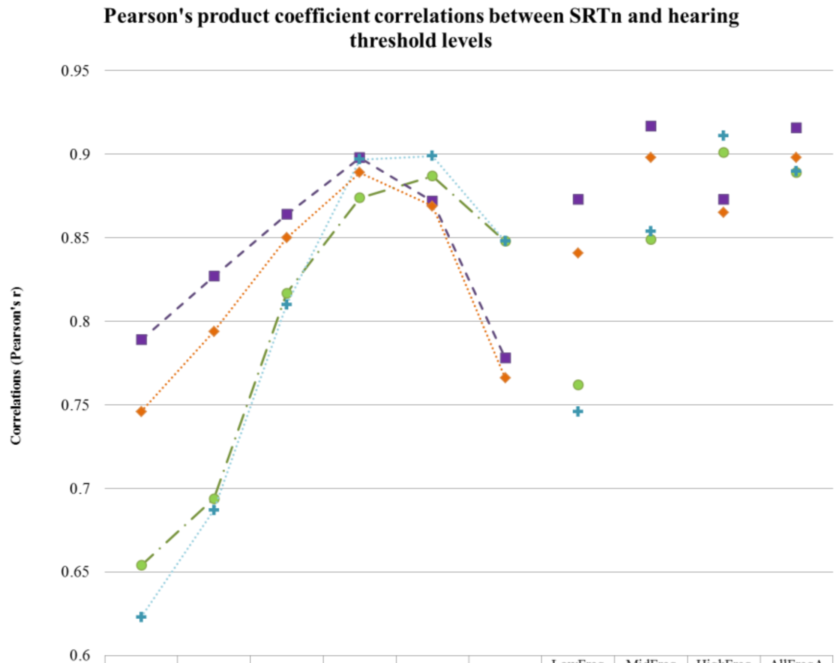


Figure 86: Boxplot showing distribution of SRTn in normal and impaired hearing particpants in all four test conditions..

Pearson product-moment correlation coefficients were generated to investigate the relationship between hearing threshold levels and the SRTn for all participants. When compared to a single audiometric test frequency, Pearson's r correlations were found to be highest at 2000 Hz or 4000Hz for all four test conditions ($r_{headphoneTSN}$ = 0.887 at

4000 Hz, $r_{telephoneTSN}$ = 0.898 at 2000 Hz, $r_{headphoneSTG}$ = 0.899 at 4000Hz, $r_{telephoneTG}$ = 0.889 at 2000Hz, $p < 0.00$). When headphones were used, highest correlation was found in the high frequency hearing threshold averages of 2000, 4000 and 8000 Hz ($r_{headphoneTSN}$ = 0.901, $r_{headphoneSTG}$ = 0.911, $p < 0.00$). Using telephone, highest correlation were found in the average of all octave frequencies from 250 to 8000 Hz and mid frequency hearing threshold averages of 500, 1000 and 2000 Hz ($r_{telephoneTSN}$ = 0.917, $r_{telephoneTG}$ = 0.898, $p < 0.00$). Figure 87 below illustrates the relationship between SRTn in all four test conditions and various audiometric combinations.



	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	LowFreq Average	MidFreq Average	HighFreq Average	AllFreqA verage
DTTCHP	0.654	0.694	0.817	0.874	0.887	0.848	0.762	0.849	0.901	0.889
DTTCPH	0.789	0.827	0.864	0.898	0.872	0.778	0.873	0.917	0.873	0.916
DTTSTGHP	0.623	0.687	0.81	0.897	0.899	0.848	0.746	0.854	0.911	0.89
DTTSTGPH	0.746	0.794	0.85	0.889	0.869	0.766	0.841	0.898	0.865	0.898

Figure 87: Correlation between SRTn in MDTT and pure tone audiometry thresholds. (LowFreqAverage = average HTL at 250, 500, 1000 Hz, MidFreqAverage = average HTL at 500, 1000 & 2000 Hz, HighFreqAverage = average HTL at 2000, 4000, 8000 Hz, AllFreqAverage = Average HTL at all octave frequencies between 250 to 8000 Hz).

Figure 88: Distribution of SRT and PTA of MDTT in test specific noise using headphone (top left) and telephone (top right) and spectreal-temporal gap noise in

headphone (bottom left) and telephone (bottom right).shows the distribution of SRTn in MDTT and average hearing thresholds at all octave frequencies between 250 to 8000 Hz for all participants. For the test using telephones in TSN, three participants reached the ceiling score of +12 dB SNR whereas five participants reached the ceiling score for the telephone in STG noise condition.

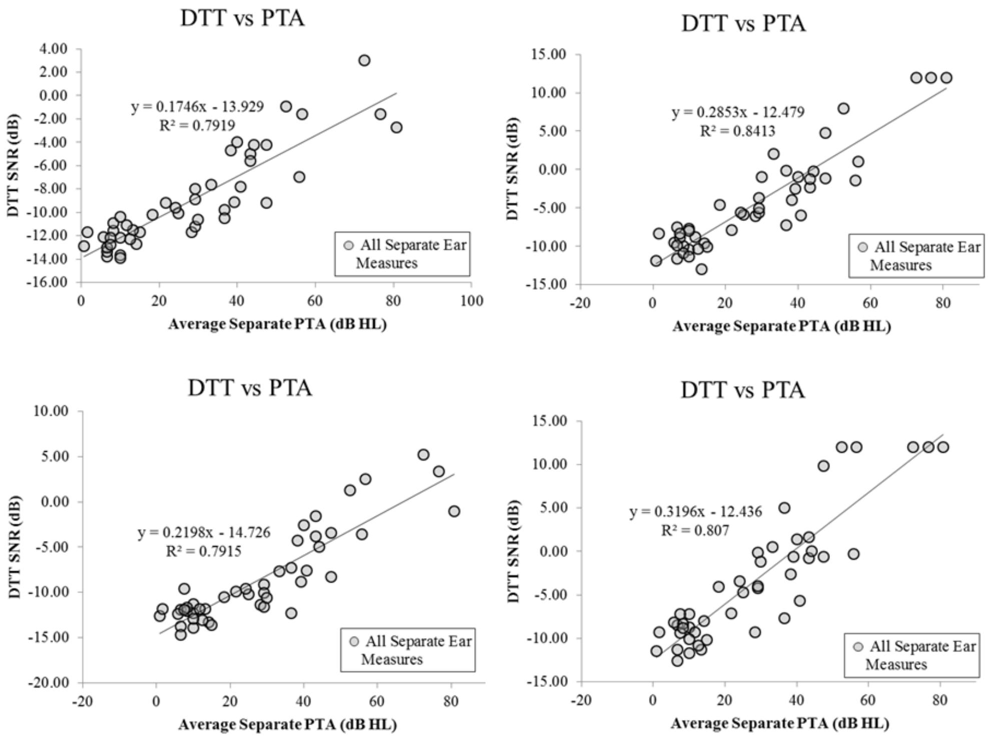


Figure 88: Distribution of SRT and PTA of MDTT in test specific noise using headphone (top left) and telephone (top right) and spectreal-temporal gap noise in headphone (bottom left) and telephone (bottom right).

One way analyses of covariance (ANCOVA) were conducted to investigate the effect of hearing loss on the SRTn in MDTT by controlling for age as the covariate. Preliminary analyses were done to evaluate the homogeneity of regression assumptions and it was found that the dependent variable (hearing levels) did not differ significantly as a function of independent variable for all four test conditions, $F_{HeadphoneTSN}$ (1, 42) = 0.693, p = 0.41; $F_{HeadphoneSTG}$ (1, 42) = 0.963, p = 0.332; $F_{TelephoneTSN}$ (1, 42) = 0.534, p = 0.469; $F_{TelephoneSTG}$ (1, 42) = 0.271, p = 0.61. ANCOVA revealed a significant effect of hearing levels on the SRTn of MDTT in all four conditions after controlling for listeners' age where, $F_{HeadphoneTSN}$ (1, 43) = 4.01; $F_{HeadphoneSTG}$ (1, 43) = 3.51; $F_{TelephoneTSN}$ (1, 43) = 12.92; $F_{TelephoneSTG}$ (1, 43) = 10.94, $p < 0.05$. Post hoc test with Bonferroni correction revealed that normal hearing listeners performed significantly better (lower SRTn) than hearing impaired listeners in all four test conditions.

Figure 89 shows changes in SRTn for individual listeners in both transducers from TSN to STG noise. Upon closer inspection, a pattern can be observed where no large changes are seen within the normal hearing group whereas hearing impaired listeners with average hearing thresholds (all octave frequencies between 250 and 8000 Hz) of more than 30 dB HL exhibited movements to the right indicating poorer performance in STG in both headphone and telephone applications. This suggests that the spectrotemporal modification does have an effect to listeners with moderate hearing losses.

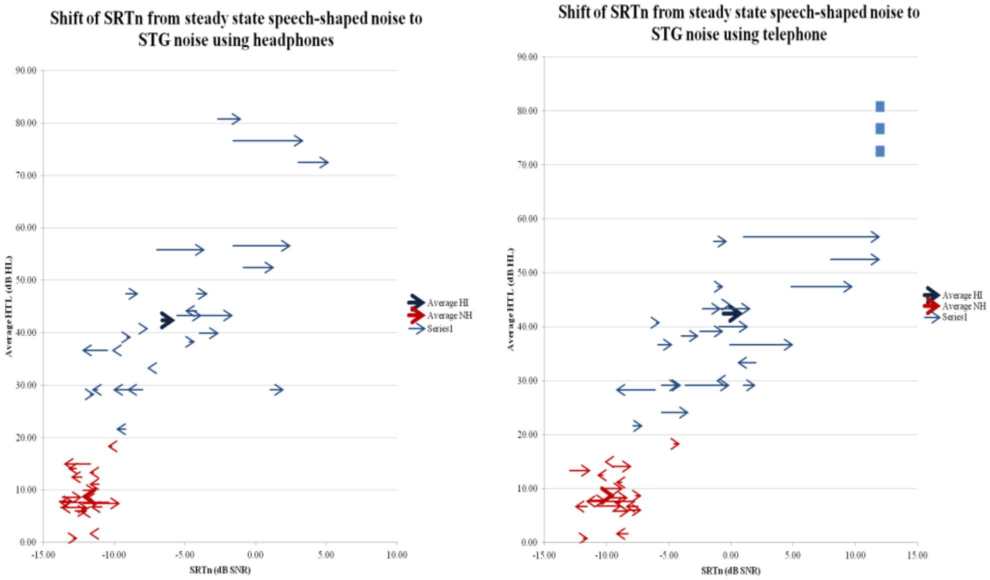


Figure 89: Shift of responses in SRTn for all listeners from TSN to STG for both headphone and telephone applications. In the test using telephones, 3 listeners with highest hearing threshold average did not show any movement of SRTn results as the score have reached ceiling score limit that was set at 12 dB SNR.

Linear regression analysis showed that an intersection at 17.32 dB HL is seen based on the linear equation for headphone testing in TSN and STG given the relationship for TSN use by; $SRT_{headphoneTSN} = 0.214*PTA - 14.608$ and for STG, $SRT_{headphoneSTG} = 0.1726 * PTA – 13.832$). The intersection found in telephone use is lower at -1.52 dB HL given the relationship for TSN; $SRT_{telephoneTSN} = 0.316*PTA - 12.261$ and $SRT_{telephoneSTG} = 0.2817*PTA - 12.313$). The intersection provides some information at which point of hearing that the performance in SRTn starts to deviate from each other when STG was used compared to TSN.

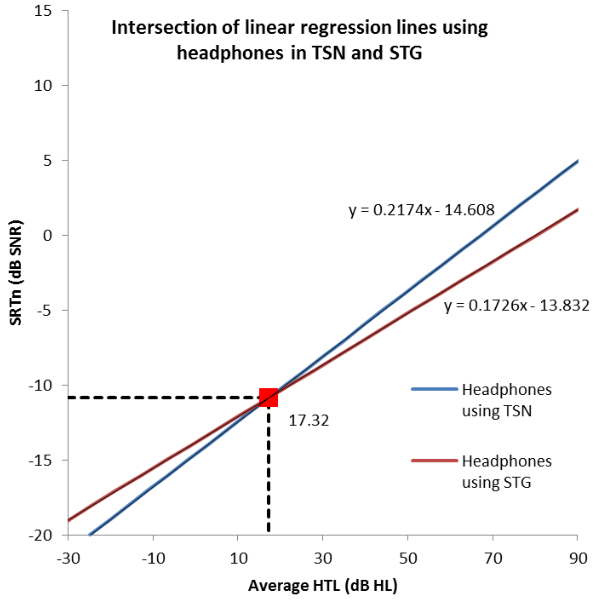


Figure 90: Linear regression lines and intersection for headphone applications in test specific noise and spectra-temporal gap noise.

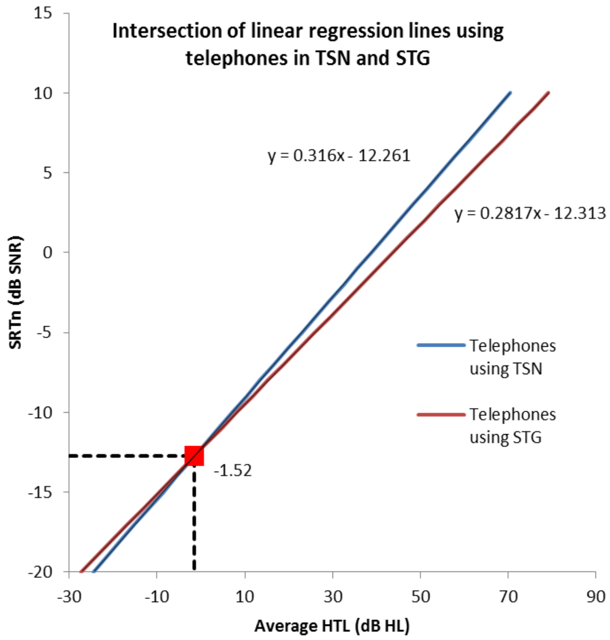


Figure 91: Linear regression lines and intersection for telephone applications in test specific noise and spectra-temporal gap noise.

To investigate differences between using TSN and STG in both types of transducers in participants with and without hearing impairment, a paired sample T test was conducted. It showed no significant differences between TSN and STG when headphones were used ($t (45) = -1.892$, $p = 0.065$). However, a statistical significance was observable when telephones were used between TSN and STG ($t(45)= -2.947$, $p < 0.05$) where the average SRTn in STG is higher at -3.48 ± 7.3 dB SNR compared to the test in TSN where the average SRTn was -4.49 ± 6.3 dB SNR.

To identify the cut-off value to estimate the true positive rate (sensitivity) and false positive rate (1 – specificity) for all four test conditions, ROC curves were generated for all four test conditions (Figure 89). Using cut-off rate of -11 dB SNR for normal hearing for the headphones in TSN presented monaurally, the test sensitivity was found the 92.3% and specificity of 85%. Using the same transducer but in STG noise and cut-off limit of -11.2 dB SNR, the sensitivity dropped to 88.5% however the specificity improved to 90%. Overall, using the telephone as transducer improved both sensitivity and specificity in both types of background noise at the cut-off limit for ‘good/pass’ hearing status at -7.6 dB SNR. Show below are the ROC analyses and tables of cut-off limits for ‘good/pass’ and ‘poor/fail’ hearing status. Figure 93 to 96 shows scatterplots of SRTn in all listeners in all four test conditions.

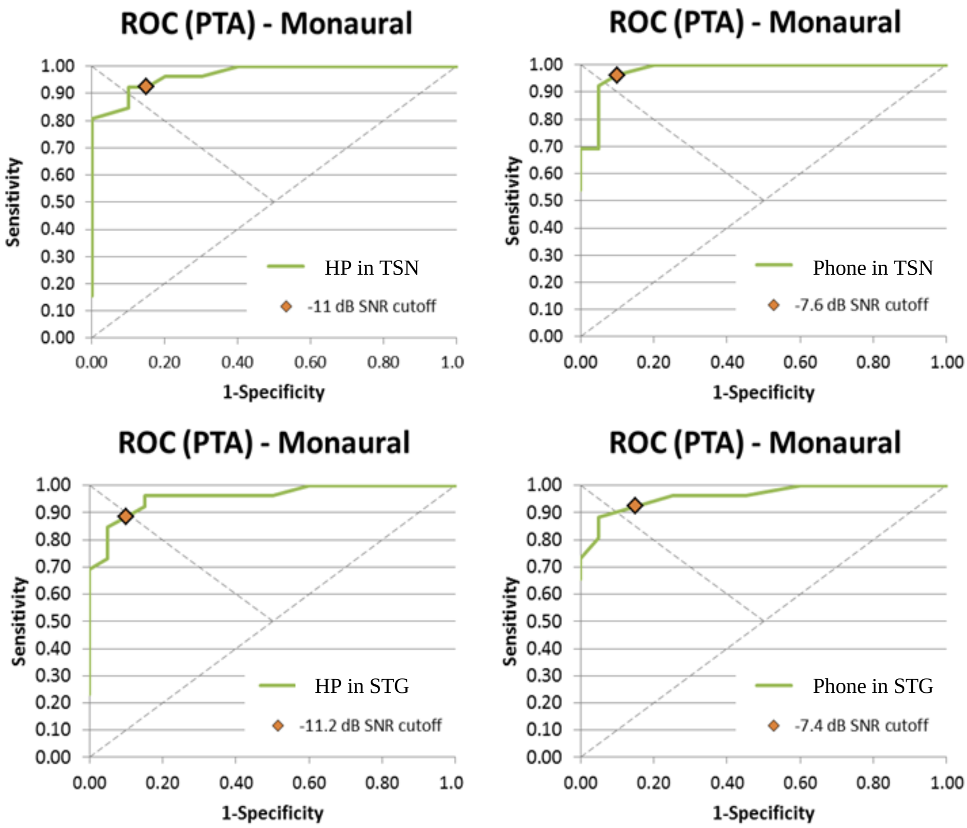


Figure 92: Receiver-operator curves for the MalayDTT in test specific noise using headphone (top left) and telephone (top right) and in spectrotemporal gap noise using headphones (bottom left) and telephone (bottom right).

Table 50: Summary of cutoff values, sensitivity and specificity for ‘good/pass’ hearing status for the MalayDTT.

	Headphone in TSN	Headphone in STG	Telephone in TSN	Telephone in STG
Cut-off NH:	-11	-11.2	-7.6	-7.6
True Positive	24	23	25	24
False Positive	3	2	2	3
True Negative	17	18	18	17
False Negative	2	3	1	2
Count	46	46	46	46
Sensitivity	92.3%	88.5%	96.2%	92.3%
Specificity	85.0%	90.0%	90.0%	85.0%
1-Specificity	0.15	0.1	0.1	0.15
Sum	1.77	1.78	1.86	1.77

Table 51: Summary of cutoff values, sensitivity and specificity for ‘poor/fail’ hearing status for the MDTT.

	Headphone in TSN	Headphone in STG	Telephone in TSN	Telephone in STG
Cut-off HI:	-9.8	-9.8	-5.8	-5.2
True Positive	21	19	21	22
False Positive	0	1	1	1
True Negative	20	19	19	19
False Negative	5	7	5	4
Count	46	46	46	46
Sensitivity	80.8%	73.1%	80.8%	84.6%
Specificity	100.0%	95.0%	95.0%	95.0%
1-Specificity	0	0.05	0.05	0.05
Sum	1.81	1.68	1.76	1.80

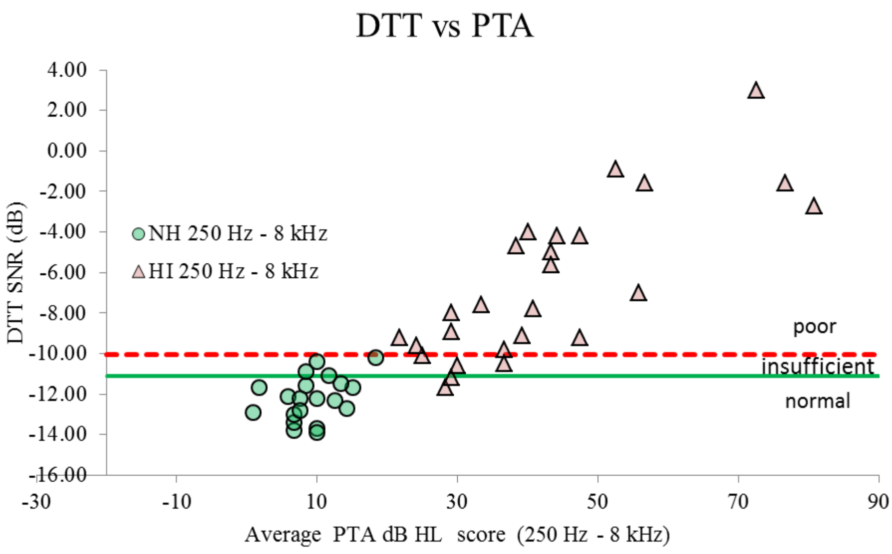


Figure 93: Scatterplot with cutoff values for both ‘good/pass’ hearing and ‘poor/fail’ hearing using headphone in test specific noise to the specific PTA for the average from 250 Hz to 8 kHz (dB HL).

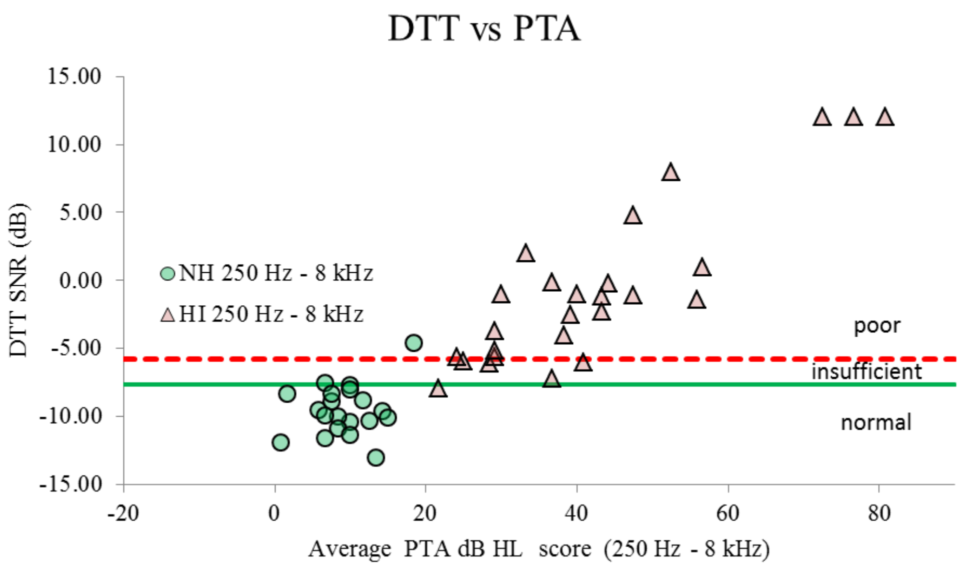


Figure 94: Scatterplot with cutoff values for both ‘good/pass’ hearing and ‘poor/fail’ hearing using telephone in test specific noise to the specific PTA for the average from 250 Hz to 8 kHz (dB HL).

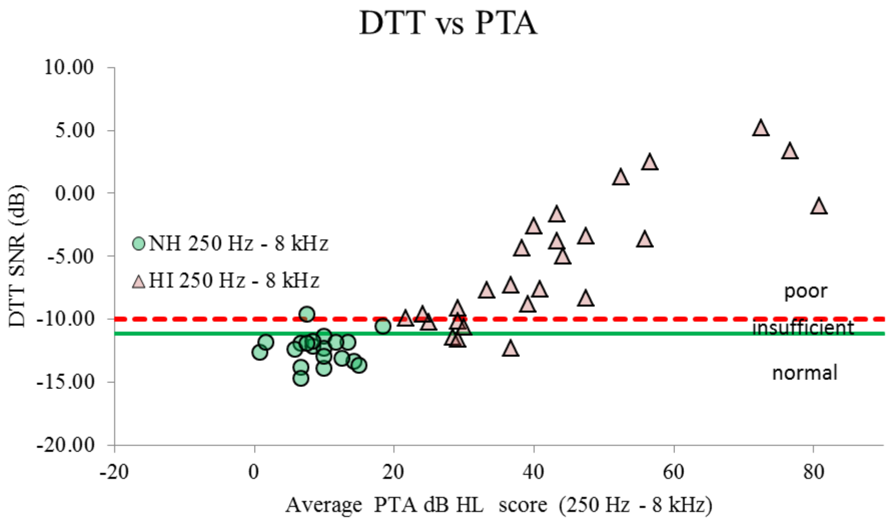


Figure 95: Scatterplot with cutoff values for both ‘good/pass’ hearing and ‘poor/fail’ hearing using headphone in spectrotemporal gap noise to the specific PTA for the average from 250 Hz to 8 kHz (dB HL).

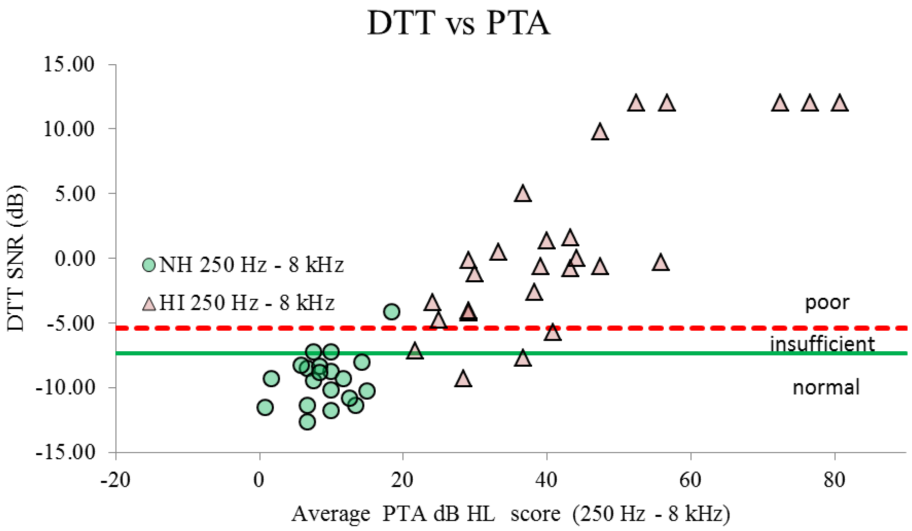


Figure 96: Scatterplot with cutoff values for both ‘good/pass’ hearing and ‘poor/fail’ hearing using telephone in spectrotemporal gap noise to the specific PTA for the average from 250 Hz to 8 kHz (dB HL).

7.1.4 Discussion

The normal hearing group obtained an average SRTn at -12.2 dB SNR for headphones in TSN and -12.4 dB SNR for headphones in STG which is very close to the normative data for adaptive measurement in Chapter 6 (where the SRTn for headphones in TSN was -12.4 dB SNR and for headphone in STG was -12.6 dB SNR). This is also true for the telephone application where the normal hearing group in this study showed SRTn in TSN at -9.5 dB SNR and in STG at -9.3 dB SNR whereas the normative values in the previous study showed SRTn using telephone in TSN at -9.4 dB SNR and in STG at -9.2 dB SNR. This again shows the remarkable consistency displayed by the MDTT. It is noted that the test conditions were controlled and performed in an audiometric cabin, therefore limiting effects that could be caused by external factors such as noise or poor audio transmission which is seen in real landline telephones (Smits, Kramer, et al., 2006). In future study, further investigation should be conducted to examine the effects of different transducers and real-world test environment for the MDTT.

Hearing impaired listeners performed significantly poorer than the normal hearing group in all four test conditions. The relationship between SRTn and hearing loss was examined and it revealed significantly strong correlations between SRTn and hearing thresholds across all test conditions. The pattern of correlation is similar to the American DTT (Figure 6 in Williams-Sanchez et al. (2014)) where correlation was strongest at 2000 Hz and drops at higher frequencies. It is also similar in terms of comparison between SRTn and average hearing thresholds at 3 or 4 octave frequencies. The correlation values in the American version is generally lower compared to the MDTT because the study was done in real-world testing environment and it was reported that very few normal hearing participants volunteered for the test. In the MDTT, using headphones was found to be highly correlated to high frequency or overall averages of the audiogram. The SRTn reported using telephones showed higher correlations either at mid frequency averages or overall averages of the audiogram. This reflects the frequency response bandwidth and audio quality of different transducers where headphones are known to have better reproducibility and wider frequency response.

The use of STG noise showed no statistical significant improvements or deteriorations in SRTn when headphones were used. This could be contributed to the small differences between the SRTn in both noise with the same transducer. Figure 89 shows the spread between TSN and STG is only visible for listener with average hearing thresholds of more than 30 dB HL and no clear movements was seen within the normal hearing groups and hearing impaired listeners with average hearing loss of less than 30 dB HL. The point of deviation between listeners where results showed the beginning of SRTn deterioration when STG noise were used are mathematically identified at 14.6 dB HL for the headphone application and -1.52 dB HL for the telephone application. Based on the linear regressions of the tests using headphones and telephones, slope of the linear regressions when TSN were used are higher when compared to BN. This suggests that performing MDTT tests in TSN regardless of the type of tranducer would be more sensitive towards hearing levels in TSN. This could also be due to the higher variablility of results obtained from hearing impaired listeners in STG.

By using 20 dB HL as the reference limit for normal hearing participants the cut-off for all four test conditions were identified. This reference value is similar to the American and Dutch version to identify the optimum criteria for the 'good/pass' hearing status criteria. The cut-off value for 'pass/good' in the MDTT is lower compare to the American English, Dutch and French DTT at -11 dB SNR for headphone in TSN and -11.2 dB SNR for headphone in STG. It is however similar to the New Zealand English version where the cut-off for 'pass' was -10.4 dB SNR and -8.65 dB SNR for the 'poor' hearing status for the monaural headphone presentation (King, 2010). The result for the STG noise could not be compared to American, Dutch and French versions as it contains different spectral properties. The differences are most likely contributed by a few factors. One reasonable explanation is that differences may have contributed by the talker-specific difference as seen in the matrix sentence study (Hochmuth, Jürgens, et al., 2015). Other possible explanation includes the fact that the data obtained in this study was exclusively obtained under controlled environment with specific transducers. Additionally, the Malay version included all disyllabic words which may have increased lexical redundancy hence making it easier for listeners to identify the digit triplets.

The use of STG improved test sensitivity and specificity for headphone use for the ‘good/pass’ limits but not in other test conditions even in the second border for ‘poor/fail’ criteria. This is also reflected in the statistical analysis. As a result, the MDTT should apply the test specific noise for its public application as no substantial benefit was observed by using the STG. The STG however showed promise in spreading the difference in SRTn for listeners with moderate hearing losses. This suggests that hearing impaired listeners with mild hearing loss or good low frequency hearing loss were able to benefit from the release in masking hence no improvements was noted for this group. For future investigation, a low pass filter can be added to the STG to reduce this effect. Using low pass filter (LPF) as background noise was found to be better at discriminating listeners with noise induced hearing loss for the Dutch Earcheck screening test as compared to a broadband speech-shaped noise or modulated speech noises (Leensen, de Laat, Snik, & Dreschler, 2011). Using LPF noise provides energetic masking to listening frequencies of up to 2000 Hz and allows release from masking from this frequency onwards. Normal hearing listeners could benefit from the release from masking and improve SRTn score when compared to the broadband noise masker. Listeners with high frequency hearing impairment will find difficulty listening to the stimuli as the low frequency information is masked hence effectively worsen SRTn scores.

7.2 Investigation of the Malay auditory-visual matrix sentence test in normal and hearing impaired listeners.

7.2.1 Introduction

Speech perception test as part of audiological test battery provides the audiologists' information to that is directly related to some of the communication difficulties that are faced by listeners. It is also known that hearing in noise is the most difficult listening task for a person with sensorineural hearing loss (Kramer et al., 1998). Information from speech tests is used either to investigate further the cause of hearing problems or help make informed decisions on hearing aid selection or adjustments. As part of routine assessments in audiology clinics, the audiometric results are commonly used to predict and compare speech reception thresholds in speech tests. However is known that speech perception performance in noise is independent from the tone audiogram (Glasberg & Moore, 1989; Killion & Niquette, 2000), nevertheless the relationship between SRTn and average audiogram values are often compared to understand relative impairment between the two tests. The SRTn obtained in the French MST in listeners with and without hearing loss were found to have significant moderately strong positive linear relationship to the pure tone audiogram average of 500, 1000, 2000 and 4000 Hz (Jansen et al., 2012).To further assess the relationship between the SRTn and PTA in the German MST, 315 ears with varying hearing loss were examined and were found to show two separate correlations between two groups of hearing levels (Wardenga et al., 2015). As a recommendation in this study, the German matrix sentence test presented in fixed background noise at 65 dB SPL may only be applicable for listeners with average hearing thresholds of less than 47 dB HL. This is because the masker may not be perceivable at 65dB SPL for listeners with hearing threshold average of more than > 47 dB HL. The effect is observed in their regression analysis as SRTn levels deviated and increased sharply from this point of hearing threshold average.

The aim of this study is to validate the MMST-AV by investigating responses obtained from listeners with varying levels of hearing in auditory only, visual only and auditory-visual modes. Listeners were asked to perform the MMST-AV in auditory-only and auditory-visual modes in test specific noise or test specific noise (TSN) and

6-talker babble noise (BN), whereas the lipreading task (visual mode) was performed in quiet. The SRTn and steepness of slope between listeners will be compared and discussed. Additionally the results of the MMST-AV in TSN will be compared to the MDTT in TSN using headphone to investigate the relationship between the two tests.

7.2.2 Methods

7.2.2.1 Recruitment of participants

The same participants who volunteered for the MDTT validation were involved in this study.

7.2.2.2 Test procedure

7.2.2.2.1 Test procedure I: Evaluation in auditory-only mode

This experiment was conducted at the IIUM Hearing & Speech Clinic, International Islamic University Malaysia in Kuantan campus. Participants were tested in a double walled audiometric cabin and were seated in front of a computer pre-installed with the UCAST software platform using normalised and equally intelligible lists. They were asked to respond by clicking word options presented in a virtual interface using a computer mouse. Before any measurements were taken, all participants completed a two training test list for each noise condition (total of four test lists of 80 sentences) to reduce training effects for both masking noise types. Participants were presented with one randomized list each for the test in TSN and BN. Sound was presented in auditory-only mode to one ear at a fixed background noise of 65 dB and signal intensity was varied to achieve desired SNR. To arrive at the speech reception threshold of 50% intelligibility (SRTn) and the performance-intensity function, equation 3 was used together with a dual track adaptive procedure using varying step sizes from Brand & Kollmeier (2002). The test was delivered using Sennheiser HD 280 Pro headphone via Creative X-Fi 51 SoundBlaster sound card.

7.2.2.2.2 Test procedure II: Auditory-visual speech perception in low SNR levels using the MMST-AV – a preliminary investigation.

In this study, lipreading (V) and auditory-visual (AV) speech perception performances of listeners with varying hearing levels were measured and compared. The AV test was measured in both test specific noise (TSN) and 6-talker babble noise (BN) and the lipreading task was delivered without any interference from background noises. Participants were trained to a closed-set response method in auditory-only mode using 4 lists (80 sentences) and no training was given for the V and AV test modes. Based on individual measurements in the auditory-only study, the logistical function was used to derive the signal-to-noise ratio level at which the participant would have responded at 30% of the words correctly. This SNR level was used to measure participants' performance in low SNR levels in the AV mode. Each participant was tested in V mode first followed by the AV mode and background noise selection was alternated between listeners. Different lists were used for each test condition and lists were randomized between listeners.

7.2.3 Results and analysis

The hearing threshold information of participants can be referred to the opening description in 7.1.3 and Figure 85.

7.2.3.1 Evaluation in auditory-only mode

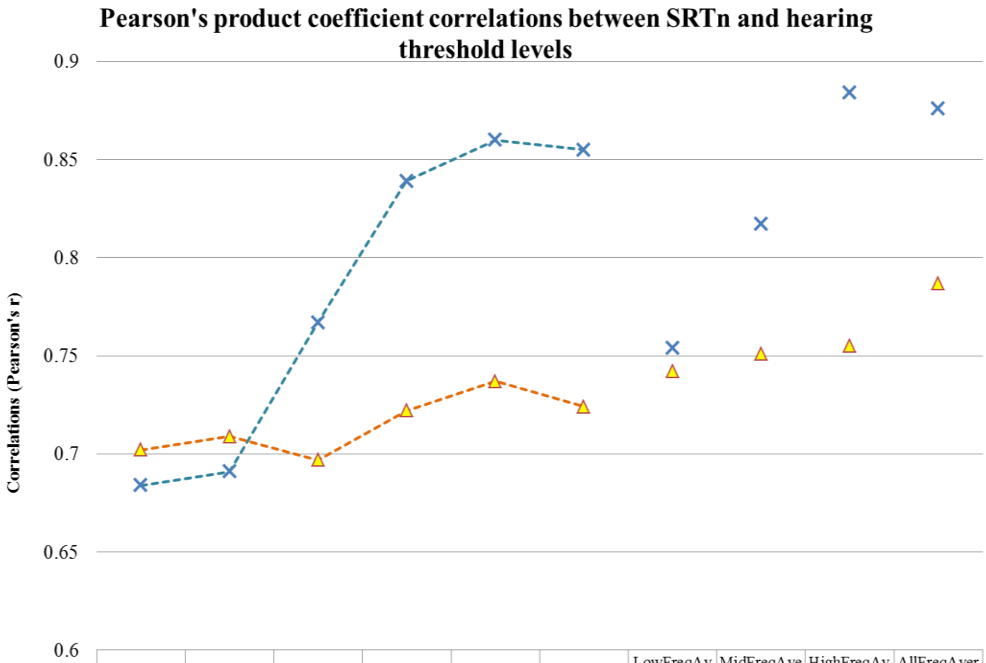
The mean SRTn and slope averaged across listeners in each group are as shown in the table below. The mean spread between normal and hearing impaired group is larger in the test in TSN compared to the test in BN at 6.5 dB compared to 5.5 dB, respectively. In terms of the differences in slope averages, normal hearing listeners show higher average slope in TSN by 2.1%/dB however the difference is only marginal for the test in BN at 0.2%/dB.

The time taken for each adaptive measurement to complete one list in order to obtain the SRTn and slope values is about 5 to 6 minutes; therefore the total time spent by participant to complete both tests in auditory-only condition was about 20 to 30 minutes including 4 sets of traning lists befire measurement. Considering the test was designed to be used repeatedly in research and clinics, the testing time is considerably long and could be improved on. One possible step is to reduce the number of sentences per list from 30 to 20 sentences only. The disadvantage to this approach is that the values for slope can not be reliably measured and only the SRTn values can be obtained (Brand & Kollmeier, 2002). A novel approach to reduce testing time by combining adaptive measurement and automatic speech recognition (ASR) to predict scores is suggested (Schädler et al., 2015). The ASR could predict individual performances during an MST measurement session and at the same time looking for significant statistical correlations for both SRTn and slope. Once an agreement is found even before test list could be completed, the test can be stopped and testing time could be reduced.

Both test in TSN and BN showed significantly positive correlation to all hearing thresholds and hearing threshold averages as shown in the figure below. The test conducted in TSN showed higher overall correlation to hearing threshold levels with highest correlation observed at the average high frequency hearing threshold (average hearing thresholds at 2000, 4000 & 8000 Hz) and overall hearing threshold levels (average hearing thresholds at all octave frequencies between 250 to 8000 Hz). The test in BN also showed highest correlation at the average high frequency hearing threshold and overall hearing threshold levels but with lower correlation coefficient levels. Figure 98 shows the SRTn of listeners in both TSN and BN compared to the overall hearing threshold average at 250, 500, 1000, 2000, 4000 and 8000 Hz. The spread between SRTn in TSN and BN seem to be uniform across the average hearing threshold levels suggesting similar effect of deterioration in SRTn caused by information masking by using the 6-talker babble noise (see Figure 97)

Table 52: Mean SRTn and slope averaged across listeners in the hearing impaired and normal hearing group in the MMST-AV in test specific noise and 6-talker babble noise.

		Test specific noise		6-talker babble noise	
		SRTn (dB SNR)	Slope (%/dB)	SRTn (dB SNR)	Slope (%/dB)
Hearing impaired group	Mean	-3.7 ± 4.2	9.5 ± 3.6	-1.8 ± 4.3	12.5 ± 3.5
	Max	8.8	18.6	15.6	22.1
	Min	-10.0	4.9	-6.4	7.4
Normal hearing group	Mean	-10.3 ± 2.5	11.6 ± 3.3	-7.4 ± 2.2	12.7 ± 4.4
	Max	-6.3	19.6	-4.0	27.5
	Min	-17.5	4.2	-12.0	6.3



	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	LowFreqAverage	MidFreqAverage	HighFreqAverage	AllFreqAverage
MSTC	0.684	0.691	0.767	0.839	0.86	0.855	0.754	0.817	0.884	0.876
MSTB	0.702	0.709	0.697	0.722	0.737	0.724	0.742	0.751	0.755	0.787

Figure 97: Correlation between Malay matrix tests and pure tone threshold levels. (LowFreqAverage = average HTL at 250, 500, 1000 Hz, MidFreqAverage = average HTL at 500, 1000 & 2000 Hz, HighFreqAverage = average HTL at 2000, 4000, 8000 Hz, AllFreqAverage = Average HTL at all octave frequencies between 250 to 8000 Hz).

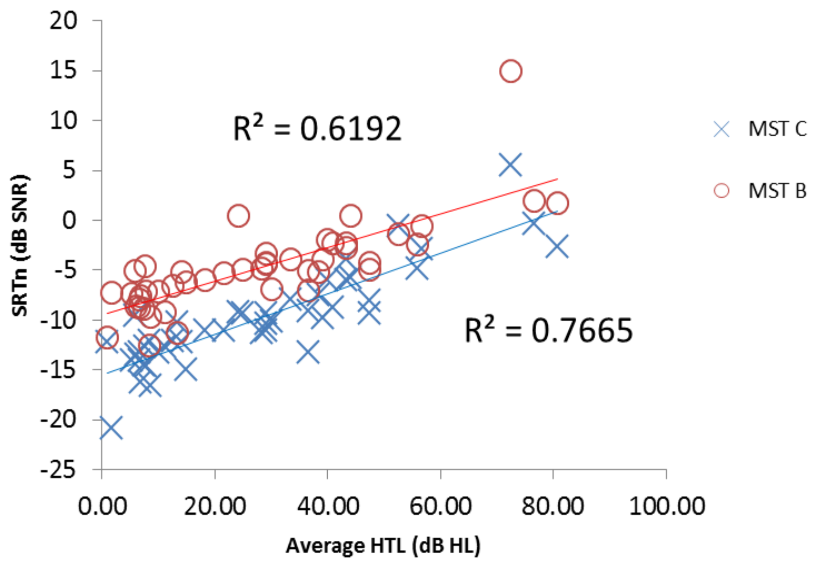


Figure 98: Relationship between MalayMST scores and average HTL at octave frequencies from 250 Hz to 8000 Hz ($p = 0.00$, Pearson correlation = 0.876 (MST C = MMST-AV in TSN); $p = 0.00$, Pearson correlation = 0.787 (MST B = MMST-AV in BN)).

One way analysis of covariance (ANCOVA) was conducted to investigate the effect of hearing loss on the SRTn in the MMST-AV by controlling for age as covariate. Preliminary analyses were done to evaluate the homogeneity of regression assumptions and the dependent variable (hearing levels) did not differ significantly as a function of independent variable for tests in TSN and BN, F_{TSN} (1, 42) = 0.01, $p = 0.98$; F_{BN} (1, 42) = 0.007, $p = 0.93$. ANCOVA revealed a significant effect of hearing levels on the SRTn of MMST-AV in TSN and BN after controlling for listeners' age, F_{TSN} (1, 43) = 4.8; F_{BN} (1, 43) =2.66, $p < 0.05$. Pairwise estimates with Bonferroni correction showed statistically significant difference ($p < 0.05$) between normal hearing listeners (estimated mean, -8.6 ± 1.0 dB SNR) compared to hearing impaired listeners (estimated mean, -5.0 ± 0.86 dB SNR). For the test in BN, pairwise estimates between normal (estimated mean, -5.7 ± 1.0 dB SNR) and hearing impaired group (estimated mean, -3.1 ± 0.86 dB SNR) showed no significant difference between groups ($p = 0.11$). The same ANCOVA analyses were conducted to investigate the effects of hearing levels on the steepness of slope in both background noises by controlling the age factor. No differences were found in the homogeneity of regression assumptions between hearing levels and slope scores in TSN and BN where, F_{TSN} (1, 42) = 7.29, p = 0.99; F_{BN} (1, 42) = 0.203, p = 0.655. ANCOVA revealed no significant effect of hearing levels on slope scores in both TSN and BN after controlling for listeners' age where, F_{TSN} (1, 43) = 0.468, p = 0.498; F_{BN} (1, 43) = 0.54, p = 0.465.

Additionally, the correlation between the MMST-AV and MDTT was investigated. Pearson's r correlation showed a significantly strong positive linear correlation between the two tests in TSN using headphones (r = 0.917, p < 0.001).

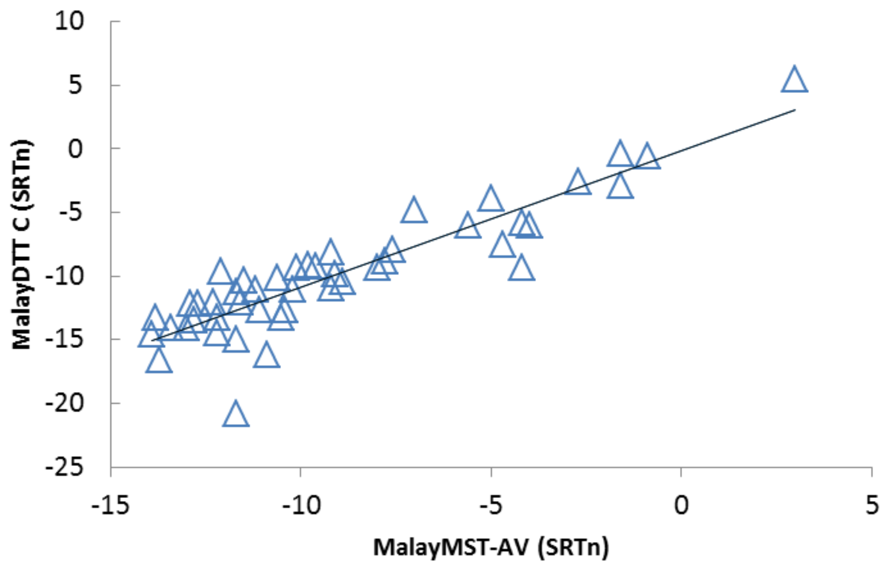


Figure 99: Relationship between SRTn score in the MalayMST-AV and MalayDTT in test specific noise using headphone (Pearson's correlation, r =0.917, p < 0.01)

7.2.3.2 Auditory-visual speech perception in low SNR levels using the MMST-AV – a preliminary investigation.

The mean percentage of visual-only score averaged across hearing impaired participants was 43.8 ± 18.5 % and 57.1 ± 17.9 % for the normal hearing listeners. Hearing impaired group performed poorer than the normal hearing listeners in both AV test conditions where speech perception score was 13.2% lower in TSN and 11.1 % lower in BN compared to normal hearing listeners. The mean, maximum and minimum scores for normal and hearing impaired listeners are shown in the table below.

Table 53: Auditory-visual and visual only scores for normal and hearing impaired listeners.

		Visual-only in quiet (%)	Auditory-visual in TSN (%)	Auditory-visual in BN (%)
Hearing impaired group	Mean	43.8	52.8	59.3
	σ	18.5	15.8	12.5
	Max	76	88.	100
	Min	16	25	38
Normal hearing group	Mean	57.1	66.0	70.4
	σ	17.9	12.9	13.5
	Max	84	86	94
	Min	25	45	44

Normal hearing listeners showed higher improvement in speech perception from auditory-only to auditory-visual where in TSN, the gain in speech perception for normal hearing listeners was 36 ± 12.5 % whereas for hearing impaired listeners gain the average gain was 22.8 ± 15.8 %. Similar pattern was observed for the gain in speech perception in BN. Gain for both participant groups were statistically higher in BN (mean, 34.83 ± 14) compared to TSN (29.4 ± 15.8) where $t\ (39) = -2.322$, $p < 0.05$.

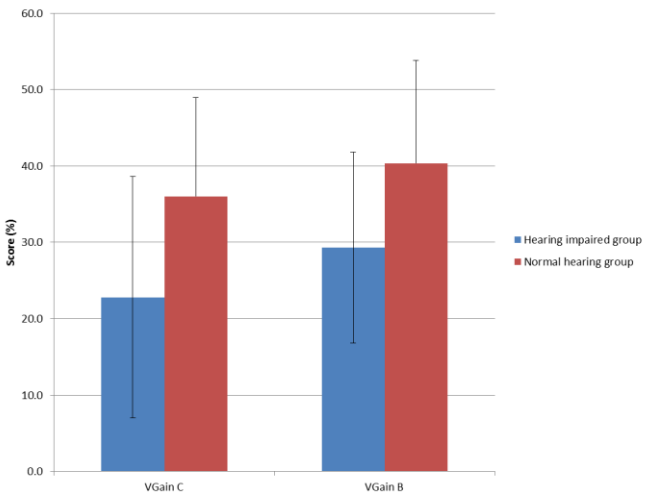


Figure 100: Percentage of improvement of speech perception scores from auditory-only to auditory-visual scores. VGainC indicates improvement in score in TSN and VGainB indicates improvement in speech perception score in BN.

One way analyses of covariance (ANCOVA) were conducted to investigate the effect of hearing loss on the AV and V performances in the MMST-AV by controlling for age as covariate. Preliminary analyses were done to evaluate the homogeneity of regression assumptions and the dependent variable (hearing levels) did not differ significantly as a function of independent variable for tests in AV_{TSN}, AV_{BN} and V only, F_{TSN} (1,5) = 1.546, p = 0.335; F_{BN} (1, 5) = 1.739, p = 0.281; F_{vonly} (1,5) = 2.39, p = 0.167. ANCOVA revealed no significant effect of hearing levels on the SRTn of MMST-AV in all AV and V modes after controlling for listeners' age, F_{TSN} (1,5) = 1.392, p = 0.387; F_{BN} (1, 5) = 1.76, p = 0.129; F_{vonly} (1,5) = 1.672, p = 0.299.

7.2.4 Discussion

As previously mentioned in the previous section, there was a significant age difference between the hearing impaired and normal hearing groups in this study. In previously published studies, age has been found to have an effect on speech perception even when the hearing impairment is mild (Dubno et al., 1984) as well as in auditory-visual tasks (Tye-Murray et al., 2008, 2007c). It is with this consideration that all of the inferential analyses were conducted by controlling the age of listeners as factor.

As expected, SRTn in listeners with normal hearing were better (lower SRTn) in both types of background noise where differences between average SRTn scores between groups were higher in TSN at 6.7 dB SNR compared to 5.6 dB SNR in BN. The normal hearing groups showed slightly steeper slope in TSN but no difference was found in BN. Both group of listeners performed lower in BN compared to TSN. This is consistent with previous MST findings using multispeaker babble noise (Hochmuth, Kollmeier, et al., 2015; Wagener & Brand, 2005) where the addition of multiple speech stimuli could have caused some confusion in language processing and increased attentional load which resulted in the deterioration of scores. In terms of evaluating both types of maskers, the findings are suggestive of TSN as the optimal masker and the reliability of BN as an informational masker.

The MMST-AV in TSN correlated well with the average hearing threshold of high frequency hearing and overall average of inter octave audiometric values. In BN, correlation coefficients were lower however it showed moderate and significant correlation with hearing levels. The correlation coefficient values for the test in TSN(r = 0.88) is very similar to the French MST's correlation (Jansen et al., 2012) to subjects with varying levels of hearing in the same type of noise (r = 0.84). Closer inspection of the average hearing levels in this study and the French MST showed similar distribution in hearing levels within listeners. Most listeners had normal to moderate hearing levels and only 3 and 2 listeners with average hearing levels above 65 dB HL in the MMST-AV and French MST, respectively. A more comprehensive evaluation of SRTn and hearing thresholds using the OLSA was conducted where 315 ears with equal distribution of hearing levels across the audiometric region were studied (Wardenga et al., 2015). The study showed two linear regressions in SRTn correlated with hearing levels in separate regions with an intersection at 47 dB HL. As the test was performed with the background noise fixed at 65 dB SPL, listeners with hearing thresholds average exceeding this intersection were assumed to be unaffected by the masking noise. This effect was not seen in this study due to its small sample size above the average hearing threshold levels of 65 dB HL.

Using the information from same 46 listeners in the MDTT and MMST-AV, the relationship between both tests were found to have significantly strong positive correlations in TSN using headphones (r = 0.917). This suggests that both tests are complementary to each other and serve as consistent tools in both hearing screening and diagnostics. This also indicates the possibility of using the MDTT as an additional diagnostic tool in assessing speech perception in noise. Using MDTT as a diagnostic tool could serve as an alternative test that is less challenging linguistically and could possibly reduce training effects seen in the MMST-AV. It could be also applied to non-native speakers of Malay with minimal effects on listeners' performance compared native speakers.

In the auditory-visual investigation, the auditory-only performance of each listener was predicted based on values obtained via dual track adaptive approach that converged at the point of SRTn. Using this method, actual measurement of speech perception performance at the level of 30% was not made hence the data shown here could vary should a measurement is made. This has limited the possibility of conducting further analysis of listeners' performances in other parameters of AV speech perception such as the visual and auditory enhancements as in Tye-Murray et al. (2007a) or integration abilities which is found to be separate from unimodal processing of speech perception (Grant, 2002). Despite this limitation, this study has shown the improvement of speech perception scores were uniform across all listeners without any discrimination against hearing levels which is consistent with previous findings (Bernstein, Auer, & Takayanagi, 2004; Grant et al., 1998; Tye-Murray et al., 2008). Listeners have shown improvements of up to 40% from auditory-only to auditory-visual speech perception modes, however the improvements from visual-only to auditory-visual conditions were not very different, which is possibly due to participants may not have been getting adequate auditory information from the auditory signal which could not be validated since actual measurement was not done. The normal hearing group performed significantly better than the hearing impaired group for V only and AV conditions which is an unexpected direction compared to the findings in Tye-Murray et al.,(2007a) as it was thought that hearing loss would improve lipreading because of practice effect. However the differences between the two groups was not more than 15 % and again, taking in consideration the predicted A-only scores, these finding may require further investigation. The improvement in BN was also found to significantly better than in TSN which could be possible as informational masking benefit more as visual aid supports audio signal.

In summary, the test in AV mode shows a great potential as a validated auditory-visual test that uses sentences which are more representative of real-world AV speech perception. The test did not show any floor or ceiling effects in any test conditions and due to its adaptive nature, and it is able to be used repeatedly for diagnostic, research or rehabilitative purposes. Improvements of test approach should be further investigated to find the optimal way to present the test in a clinical setting. For future investigations, the training effects and list equivalency in AV mode could be investigated as well as incorporating other useful auditory-visual recording paramaters to monitor individual AV speech perception performance, such as specific abilities for each listener to utilize both unimodal conditions and ability to integrate both modalities as a predictor to their AV abilities.

CHAPTER 8

CONCLUSIONS

8.1 Development of tests

8.1.1 Speaker selection

Based on previous literatures, listener's performance in speech perception tests are highly dependant on the speaker or talker characteristics (Bradlow et al., 1996; Hochmuth, Jürgens, et al., 2015; Mullennix et al., 1989; Versfeld et al., 2000) therefore the study was conducted to ensure the best possible candidate to provide her voice and video for both MDTT and MMST-AV. The selected speaker had similar rate of speech at 3.6 syllables per second compared to other MSTs which is between 3.4 syllables per second for the Russian MST (Warzybok, Zokoll, et al., 2015) and up to 5.5 syllable per second for Spanish MST (Hochmuth et al., 2012). Due to careful selection of the speaker, no floor effect was observed in the lipreading (V only) task in normal and hearing impaired listeners.

8.1.2 Digit selection for MDTT

Disyllabic Malay digits zero (kosong), one (satu), two (dua), three (tiga), four (empat), five (lima), six (enam), seven (tujuh) and eight (lapan) were selected for this test. The trisyllabic digit nine or 'sembilan' was omitted to maintain digit homogeneity. The carrier phrase 'nombor' or number was chosen instead of 'digit' or the digits because the term is more commonly used in Malaysia.

8.1.3 Recording of digits, generating masking noises and testing apparatus for the MDTT

Two takes of digit recordings were used for the normalisation procedure to allow flexibility in choosing the digits with the steepest slope of intelligibility. Two types of background noise were produced as potential masker for the MDTT which were the test specificand the spectrotemporal gap noises. The aim of generating two types of masking noise was to evaluate which would help to increase test sensitivity and specificity in screening for hearing impairment. A specific set of headphones and sound card were used throughout this study. Software for the MDTT was written by Assoc. Prof. Greg O'Beirne using the Labview (National Instruments, TX, USA).

8.1.4 Word selection for the MMST-AV

A standardized corpus for the Malay language was not available at the time the study was initiated, therefore a list of 1000 most frequent words were phonetically transcribed from a Malay newspaper daily "Utusan Malaysia" to produce a reference corpus for this study. Words that were disyllabic, highly frequent and neutral semantically were chosen from this reference corpus. The fifty chosen words were found to correlate very well with the phonemic distribution of the reference corpus.

8.1.5 Video recording, generating masking noises and testing apparatus for the MMST-AV

High definition audio and video recording of the sentences were done concurrently in an audiometric room at the Department of communication disorders research facility. A custom mechanical brace was made using plaster gauze/plaster of Paris to hold speaker's head and neck during recording to avoid movements. This proved to be an important element to ensure concatenated words would not produce video judders that were too obvious.

To present the desired sentences, the software that was developed by Assoc. Prof. Greg O'Beirne was able to generate sentences in real time by concatenating video and audio fragments on demand. Two masking noises were developed for the MMST-AV; the test specific noise was design to provide energetic masking whereas the 6-talker babble noise was generated to provide informational masking. The scoring method was adapted O'Beirne et al., (2015) as the MMST-AV used similar recording and editing methods that required a novel scoring approach.

8.1.6 Normalisation of the MDTT and constructing equivalent lists.

Twenty normal hearing listeners were recruited to assist in the normalization of the MDTT in both TSN and STG using headphone and telephone. Average level adjustment for selected digits is as shown in Table 24 on page 150. Digits were selected from Take 1 and 2 based on the steepness of their slopes. Some digits showed higher than expected (i.e artefactual) steepness and special considerations were given to ensure the data was accurately fitted using Equation 2 on page 146. Using the selected digits, unique sets of 27 digit triplets per list were assigned in 8 equally intelligible lists in each four test condition.

8.1.7 Normalisation, refinement and generating equivalent lists for the MMST-AV

Due to the unique way the MMST-AV was recorded and edited, two possible methods of normalisation were identified. The word normalisation method gave more consistent results and was chosen as the preferred method of level optimization. In terms of video refinement, video transitions that were within 1 standard deviation of the levels of judder recorded in the original recording were found to have statistically indistinguishable judders when compared to the original recording, and this informed the selection of the lists

Word occurrences between lists were not balanced and many considerations were taken to ensure all sentences were unique across lists and between tests. However, all the lists showed similar balance in terms of its phoneme distribution. In the subsequent study, it was found that the disproportionate word distribution did not affect the average SRTn and slope scores.

To be able to produce reliable slope and SRTn values, the MMST-AV was designed to include 30 sentences in each list as recommended by Brand & Kollmeier (2002). Due to this design, the length of testing for each measurement was about 5 to 6 minutes which is long considering that a key feature of the MST design was to be able to conduct the measurements repeatedly. Other MST versions in the Hearcom project use 20 sentences to complete one set of measurement. However, this smaller number of sentences is only possible by adaptively tracking the 50% target alone (and hence estimating the SRT only) rather than tracking the 20% and 80% targets (which permits concurrent estimation of both SRT and slope value). Taking into consideration the recent developments in automatic speech recognition application in speech tests (Schädler et al., 2015; Schell-Majoor, Rennies, Ewert, & Kollmeier, 2015), a predictive model could be used during measurements to assist in the accurate estimation of SRTn and slope as well as reduce testing time.

8.2 Evaluation of the digit triplet and matrix sentence test in Malay

8.2.1 Evaluation of list equivalency of MDTT

Evaluations of the MDTT lists were done using fixed signal-to-noise ratio and adaptive measurements for all four test conditions. The lists were found to be statistically equivalent in both types of measurement. The reference SRTn for normal hearing listeners by fixed SNR measurement for the MDTT are; -11.3 ± 0.34 dB SNR using headphones in TSN, -11.9 ± 0.4 dB SNR using headphones in STG, -10.24 ± 0.1 dB SNR for telephone in TSN and -10.8 ± 0.3 dB SNR using telephone in STG. The reference normative SRTn values for normal hearing listeners using adaptive measurement were lower using headphones at -12.44 ± 0.1 dB SNR for in TSN and -12.7 ± 0.3 dB SNR in BN but slightly higher when telephones were used at -9.43 ± 0.2 dB SNR in TSN and -9.24 ± 0.26 dB SNR in BN. The MDTT measured in fixed SNR levels were found to be comparable to other versions of DTT although it is the lowest compared to other versions of DTT (Jansen et al., 2010; Smits et al., 2004; Wagener et al., 2005).

8.2.2 Evaluation of list equivalency of the MMST-AV

The MMST-AV lists were also evaluated in both fixed and adaptive measurements. For the adaptive measurement, this study adopted the dual track adaptive approach as proposed by Brand & Kollmeier (2002). A total of 40 normal hearing listeners participated in the evaluation of the test. Lists were found to be statistically equivalent in terms of SRTn and slope in both types of measurement approach. Similar to other versions of MST, a significant training effect was observed in the MMST-AV especially from the first and second test lists. It is therefore recommended that two training lists are provided to listeners before any measurement using the MMST-AV is taken. The investigation of the SRTn and slope measurement (using fixed SNRs) provided the normative reference for normal hearing listeners which are; -10.1 ± 0.2 dB SNR at a slope of 14.9%/dB using TSN. For the test in BN, the normative SRTn and slope values were -6.4 ± 0.2 dB and 12.2%/dB, respectively.. The reference value in TSN for the MMST-AV share the lowest recorded SRTn with the Finnish MST (Dietz et al., 2014). . Using adaptive measures slightly reduced the SRTn values in both types of noise where in TSN SRTn was -10.4 ± 0.6 dB SNR and -6.5 ± 0.6 db SNR when BN was used.

8.3 Validation of test

8.3.1 Evaluating MDTT in listeners with varying hearing levels.

As expected, normal hearing listeners performed significantly better (lower SRTn) than their hearing impaired counterparts after the interaction between hearing thresholds and age were controlled for. A strong significant correlation was found in between SRTn and participants' hearing threshold levels especially for hearing thresholds in high frequencies and hearing thresholds averaged across all audiometric octave frequencies. This suggests that the MDTT is sensitive to hearing levels especially at high frequencies.

The cut-off applied for pass criteria using ROC analyses showed that the MDTT has sensitivity and specificity of at least 85% across all four testing conditions. However all of the tests were conducted under controlled environment and further investigation is necessary to identify factors that could affect SRTn in real-world testing environment.

The expected benefit of using STG for normal hearing listeners to be released from masking were minimal when tested in STG as no statistical difference in SRTn was detectable. However Figure 89showed that using STG in both transducers (headphone and telephones) showed some potential in spreading the SRTn differences between normal and hearing impaired listeners especially for listeners with hearing threshold average of more than 40 dB HL.

8.3.2 Evaluating MMST-AV in auditory-only, visual-only and auditory-visual modes in listeners with varying hearing levels.

Similar results were observed in listener's SRTn in auditory-only condition where normal hearing listeners showed significantly lower SRTn than hearing impaired listeners. Using 6-talker babble noise as an informational masker had indeed considerably deteriorated SRTn performance of listeners of both normal hearing and hearing impaired groups.

The MMST-AV scores from tests conducted in auditory–only mode were found to correlate well with high frequency hearing thresholds and averaged thresholds across all audiometric octave frequencies with the test in TSN showing stronger correlation than the test in BN.

In the visual-only and auditory-visual modes, it appeared that normal hearing listeners showed overall better lip reading and auditory-visual performances than hearing impaired listeners. However, after controlling for the effect of age and hearing levels of participants, no differences were found between both groups. The tests in both visual-only and auditory-visual conditions did not show any floor effect as all participants were able to score above change score.

8.4 Limitation of study

8.4.1 Software development

The software for delivering MDTT and MMST-AV using the UCAST platform - was the first of its kind to be developed outside of the Hearcom initiative in Europe. The software is able to form sentences on demand instead of using pre-synthesized sentences and more importantly it allows for auditory-visual mode of signal presentation. A considerable amount of time was spent into the development process of the software, whereby multiple testing and re-evaluations were made to prepare it for the various applications in this study. As a result, the author was unable to proceed with a more comprehensive testing on the auditory-visual part of the study due to time constraints.

8.4.2 Recording and editing sound and video files for MMST-AV.

The custom head cast was intended to ensure minimal head and neck movement during recording, yet a substantial amount of videos had to be rejected as the amount of judder produced after concatenating words were found to be significant. Additionally, the word fragment “saya_mahu” was omitted after a review by the author due to poor editing ‘out points’ (point at which the media file ends) was identified. These errors in recording and editing had limited the number of words that could be selected for the test; however adequate numbers of videos were still usable to form equally intelligible lists for the MMST-AV.

8.4.3 Issues with sample size & distribution

As the tests require a certain number of native Malay speakers as participants, certain logistical and financial limitation was incurred during this study as it was based in Christchurch, New Zealand.

For the evaluation of list equivalency for the MDTT using fixed SNT measurement, participants were tasked to complete what is essentially one SNR for each list. At this point of study, data was collected in Christchurch during the summer holiday period which led to only a small group of participants. With a total of 16 participants, ultimately only 8 sets of data points were acquired for this investigation. The data were found to be normally distributed however increasing the sample size could have improved the power of study.

To validate the MDTT and MMST-AV a group of 26 hearing impaired participants were recruited via convenience sampling method. The majority distribution of hearing threshold levels within this group was generally concentrated within the mild and moderate hearing levels. Only three participants had average hearing threshold levels of more than 65 dB HL. This has limited the observation of SRTn and slope performances of listeners with more severe hearing loss. Wardenga et al. (2015) saw a deviation of trend for listeners with hearing threshold average of more than 47 dB HL, citing the possibility that listeners would be less affected by the fixed background noise (65 dB SPL) due to issues of audibility. This observation was only possible due to the careful selection of participants with equal and adequate number of samples across hearing threshold levels. The age distribution between the normal hearing and hearing impaired participants were also significantly different hence precautions had to be made by controlling the effect of age.

8.4.4 MDTT in real-world testing environment

Although it is known that testing hearing disability using DTT was is relatively robust against types of transducer and household noise (Culling et al., 2005; Smits et al., 2004), there many other parameters that needs to be considered in real-world applications. The current study for the MDTT was only performed under laboratory environment where the type of transducer, sound card and ambient room noise were controlled. For example, in Leensen & Dreschler (2013) identified significant effect of test environment where results of the Earcheck screening test were 1 dB poorer that the SRTns obtained in laboratory environment for normal hearing participants. However, a marginal improvement of SRTn was observed in the hearing impaired groups. It was possible that participants conducted the test in rooms that were reverberant or too noisy hence affecting the test. As it is expected that using different types of transducer and sound card would not have a large effects on the DDT (Culling et al., 2005), the result of this study is likely to be applicable to test presentation in quiet environment for the time being until further investigations into the effect of noise and reverberration can be carried out.

8.4.5 Auditory-visual testing

The result presented in the auditory-visual study gave some insight on the lip reading and auditory-visual performances of listeners with varying hearing levels using the MMST-AV. It was assumed the best testing procedure was to derive the SNR levels based on the auditory-only performances of the listeners. Due to this assumption, participants' performances were not measured again in auditory-only mode following switching to the fixed SNR mode when measuring specific modality. This could have led to under or over estimation of the scores and this may have limited the study's ability to implement further analysis of individual auditory and visual processing abilities. Other testing parameters such as the training effect in AV mode, list equivalence in AV mode and auditory presentation levels were yet to be investigated due time constraints.

8.5 Future recommendations

8.5.1 Implementation of MDTT in real-world environment

As the study was fully conducted in laboratory test environment, the MDTT should be evaluated in real-world testing environment to ensure all factors are taken into consideration. As previously mentioned in the last section, there are several environmental factors during the test that could affect the result of an adaptive speech test presented in signal-to-noise ratio in a noisy and/or reverberant environment. Additionally, factors within listeners using the telephone should be fully understood since using a screening tool over the internet or telephone is a new in Malaysia. Performances of callers who differ in gender, age and especially ethnicity could vary and affect test results.

8.5.2 Improving recording methods for the auditory-visual MST

Despite the best effort to maintain video quality using the custom head cast, judders were still found after the words are put together in sentences. Additionally, the cast was conveniently hidden under the speaker's headscarf which may not be applicable should the test be replicated in other languages/cultures. Improvements can be made by designing a custom head cast using 3D scans. This allows more flexibility in designing the head cast to include structural support to the back and neck to ensure minimal head movement. 3D printing allows the cast to be design in one piece like the plaster gauze or broken into selected pieces that could allows for better support and point of reference for speaker, while simultaneuouslyconcealing it beyond the camera's view. As small movements during recording can affect the quality, it is strongly advised that the speaker is left alone during the recording session, or that the recording takes place on a concrete or similarly hard floor.

8.5.3 Improving testing time for MDTT and MMST-AV

The MDTT takes about 2-3 minutes to complete and the MMST-AV would take slightly longer - up to 10 minutes as training sessions are necessary to get a reliable measurement. Attempts should be made to make testing time shorter especially when several measurements are requiredsuch as for hearing amplification and cochlear implant candidacy evaluations.

For the MDTT, significant amount of time can be saved for normal hearing listeners by stopping the test before the allocated digit triplets is completed. One suggestion is to apply a special ‘pass’ rule should a listener perform well consistently below the cut-off SNR point for normal hearing. Another method that could possibly expedite the measurement process is to conduct a digit scoring approach instead of the current triplet scoring method. By collecting more data, statistical strength can be improved and used to calculate the scores more efficiently. This method could possibly reduce the number of triplets necessary within each list.

To improve testing time for the MMST-AV, several techniques are proposed.

i. Reduce number of sentences per list from 30 to 20, which effectively would abandon any slope values However, slope values could not be estimated using this approach.

ii. By recruiting more participants with varying hearing loss equally across the audiometric range and testing them in other MST parameters (such as presentation levels, noise type, open or closed-set response), a predictive model could be made and later be used to be compared with listener’s responses. If a statistical significant correlation is found between the model and responses given by the listener and all other testing parameters are met, the measurement could be stopped before the completion of all sentences.

iii. Finally, a closer look at the training lists study could be used to identify the point which the SRTn and/or slope of listeners could have stabilizes, instead of relying on the completion of two training lists.

8.5.4 Explore other possible masker in addition to modified STG noise to improve MDTT sensitivity

The development of the STG noise has showed potential in spreading the difference between normal and hearing impaired listeners – however, this trend was only observed in listeners with average hearing thresholds of more than 40 dB HL. A follow up study could be implemented to better understand at which region hearing would benefit from the release from masking by examining performances of listeners in various types of masking noise. It is suggested that the following noises are used based on the study for the Earcheck hearing screening test (Leensen & Dreschler, 2013) for future investigations.

i. Test specific noise as the reference noise.

ii. Low pass noise with 15 dB SPL noise floor and crossover frequency of about 2000 Hz.

iii. A high pass noise with 15 dB SPL noise floor and crossover frequency of about 2000 Hz as a control for high frequency hearing.

iv. Spectrotemporal gap noise with added low pass noise with 15 dB SPL noise floor and crossover frequency of about 2000 Hz,

v. Spectrotemporal gap noise with added high pass noise with 15 dB SPL noise floor and crossover frequency of about 2000 Hz.

Additionally, the number of participants to be recruited should be representative of hearing levels across the audiometric range for this study.

8.5.5 Evaluation of different parameters in MMST-AV

In this study the MMST-AV was testing under headphones only using a closed-set response method which was presented a fixed background noise of 65 dB SPL. As the test can be used with much more flexibility, other parameters of MMST-AV should be tested to identify the normative values or to examine the effects it could have to the result of the test. The parameters which could benefit from investigation are as shown below.

i. Adaptive evaluation of MMST-AV in open-set responses adaptively. It has been shown that the SRTn result could deteriorate as much as 1 dB between listeners (Hochmuth et al., 2012) from open to closed-set responses.

ii. Evaluating effects of varying background noise levels.

iii. Evaluation of MMST-AV in quiet.

iv. Free-field applications of MMST-AV.

8.5.6 Evaluating MMST-AV in auditory-only condition in moderate to profound hearing listeners

Due to the limitation of recruitment of participants in this study for listeners with moderate to profound hearing losses, the effects of hearing levels was not fully examined. As found in a previous study (Wardenga et al., 2015), the effects of noise was reduced in listeners with hearing threshold average of more than 47 dB HL and this effect is likely to be applicable to all MST tests. Hence further investigation is necessary by recruiting more listeners in this hearing level bracket.

8.5.7 Evaluation of the MDTT and MMST-AV in non-native Malay speakers

In Warzybok, Brand, et al. (2015) SRTn of non-native German speaker were significantly poorer thanthose of native speaker for the OLSA test. Similar investigations should be done in this context in Malaysia as it is a multi ethnic nation comprising of 63.1% ethnic Malay, 24.6% ethnic Chinese, 7.3 % ethnic Indian and 5% other ethnic minorities (Raof, 2010). Additionally, eventhough most ethnic group converse regularly in their first language, Malay language is taught formally throughout primary and secondary schools with a vast majority of the population able to converse and read in Malay. Although a previous study indicated that native and non-native listeners show similar SRTn in a closed-set digit triplet tests (Warzybok, Brand, et al., 2015), due to the unique mix of ethnic backgrounds in the Malaysian community, the effect of listener's first language should be investigated.

8.5.8 AV testing using MMST-AV

Further refinement of the AV testing modes is necessary to fully optimize this function. Several issues need to be clarified such as the training effect of the test in AV. The test involves multi sensory testing and well as auditory-visual speech and language memory assessment. This could be overwhelming at first, especially for first time listeners and thus reduce the possible score that could be obtained by an individual.

If the test is to be applied as a rehabilitation tool to train listeners to be better at AV speech perception in noise, further investigation should be done to identify best clinical approach to this apply this test. This could also indicate that a list equivalency evaluation in AV mode is necessary.

The effect of stimuli level presentation could also be studied to understand AV speech perception in low, high level of intelligibility in noise.

8.6 Final remarks

To develop and evaluate the digit triplet and the auditory-visual matrix sentence tests took a considerable amount of time and effort to ensure recording and measurement procedure could be implemented accurately.

The Malay digit triplet test showed high sensitivity and specificity in detecting hearing impairment under controlled testing conditions. The test could be applied using headphones for internet delivery, or using telephones over landlines. At the moment, using the test specific noise would be an optimal masker for the test. The test would be of great benefit to the Malaysian population to would like to seek an objective measurement of their hearing ability.

The Malay auditory-visual matrix sentence test is the first of its kind in the Malay speaking population. This test could provide immense benefit to the audiology practitioner in assessing speech perception in noise. It can be administered in a closed-set test format at a fixed 65 dB SPL background noise level, using the test specific noise and 6-talker babble noise. The test was developed according to the recommendations by ICRA and the normative values indicated in this study are comparable to other versions of matrix sentence test. The addition of an auditory-visual condition allows for more applications in terms of measurement of auditory-visual speech perception ability and may potentially serve as a tool for aural rehabilitation.

REFERENCES

Abu Bakar, A. (2007). *Malaysia Health: National ear and hearing disorders survey*.

Aibara, R., Welsh, J. T., Puria, S., & Goode, R. L. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research*, *152*, 100–109.

Akeroyd, M. A, Arlinger, S., Bentler, R. A, Boothroyd, A., Dillier, N., Dreschler, W. A., … Kollmeier, B. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests. *International Journal of Audiology*, (November 2014), 1–6. doi:10.3109/14992027.2015.1030513

ASHA. (1996). Guidelines for audiologic screening. In *ASHA Desk Reference* (Vol. 4, pp. 333–382).

Asmah, O. (1971). Standard langauge and the standardization of Malay. *Anthropological Linguistics*, *2*(Feb), 75–89.

Asmah, O. (1974). The Possessive Phrase in the Western Austronesian Languages. In *Papers of the First International Conference on Comparative Austronesian Linguistics* (Vol. 13, pp. 391–407).

Asmah, O. (1993). *Susur galur bahasa melayu.* Kuala Lumpur: Dewan Bahasa Pustaka.

Bellis, T. J., & Ferre, J. M. (1999). Multidimensional approach to the differential diagnosis of central auditory processing disorders in children. *J Am Acad Audiol*, *10*, 319–328.

Benjamin, G. (2009). Affixes , Austronesian and iconicity in Malay. *Bijdragen Tot de Taal, Land- En Volkend*, *165*(2), 291–323.

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1-4), 5–18. doi:10.1016/j.specom.2004.10.011

Bexelius, C., Honeth, L., Ekman, A., Eriksson, M., Sandin, S., Bagger-Sjöbäck, D., & Litton, J. E. (2008). Evaluation of an internet-based hearing test - Comparison with established methods for detection of hearing loss. *Journal of Medical Internet Research*, *10*. doi:10.2196/jmir.1065

Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., & Graeme, M. (1989). Speech perception using combinations of auditory , visual , and tactile information. *Journal of Speech Rehabilitation Research and Development*, *26*(1), 15–24.

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology*, *2*(1), 3–10.

Bowden, A. (2013). *Normalisation , Evaluation and Verification of the NEW ZEALAND HEARING SCREENING TEST* . University of Canterbury.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3), 255–272. doi:10.1016/S0167-6393(96)00063-5

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, *111*(6), 2801. doi:10.1121/1.1479152

Brungart, D. S., Simpson, B. D., Ericson, M. a, & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5 Pt 1), 2527–2538. doi:10.1121/1.1408946

Byrne, D., Dillon, H., Tran, K., Cox, R., Hagerman, B., Hetu, R., … Frolenkov, G. I. (1994). An international spectra comparison of long-term average speech. *The Journal of the Acoustical Society of America*, *96*(October), 2108–2120.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543. doi:10.1016/j.tics.2007.10.001

Campbell, R., & Dodd, B. (1980). Hearing by eye. *The Quarterly Journal of Experimental Psychology*, *32*(April 2014), 85–99. doi:10.1080/00335558008248235

Carhart, R. (1952). Speech audiometry in clinical evaluation. *Acta Oto-Laryngologica*, *41*(1-2), 18–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14932955

Carson, A. J. (2005). "What brings you here today?" The role of self-assessment in help-seeking for age-related hearing loss. *Journal of Aging Studies*, *19*(2), 185–200. doi:10.1016/j.jaging.2004.07.001

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*, *17*(5), 331–49. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324094&tool=pmcentrez&rendertype=abstract

Culling, J. F., Zhao, F., & Stephens, D. (2005). The viability of speech-in-noise audiometric screening using domestic audio equipment. *International Journal of Audiology*, *44*(12), 691–700. doi:10.1080/14992020500267017

Cullington, H. E., & Zeng, F.-G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, *123*(1), 450–61. doi:10.1121/1.2805617

Dallos, P., Billone, M. C., Durrant, J. D., Wang, C., & Raynor, S. (1972). Cochlear inner and outer hair cells: functional differences. *Science*, *177*(46), 356–358.

Davis, H. (1983). An active process in cochlear mechanics. *Hearing Research*, *9*(1), 79–90. doi:10.1016/0378-5955(83)90136-3

Degeest, S., Keppler, H., & Corthals, P. (2015). The Effect of Age on Listening Effort. *Journal of Speech , Language and Hearing Research*, 1–9. doi:10.1044/2015

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic Loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769–773.

Denes, P. B., & Pinson, E. N. (1993). *The speech chain: the physics and biology of spoken language* (Second edi.). New York: W. H. Freeman and Company.

Diehl, R. L., & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, *1*(2), 121–144. doi:10.1207/s15326969eco0102

Dietz, A., Buschermöhle, M., Aarnisalo, A. a, Vanhanen, A., Hyyrynen, T., Aaltonen, O., … Kollmeier, B. (2014). The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. *Acta Oto-Laryngologica*, *134*(7), 728–37. doi:10.3109/00016489.2014.898185

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293–321. doi:10.1016/j.cognition.2008.09.008

Dillon, H. (2000). *Hearing Aids* (1st ed.). Sydney: Boomerang Press.

Dillon, H., & Ching, T. Y. C. (1995). What makes a good speech test□? In *Profound Deafness and Speech Communication* (pp. 305–344).

Dreschler, W. A., & Plomp, R. (1980). Relation between psychophysical data and speech perception for hearing-impaired subjects. I. *The Journal of the Acoustical Society of America*, *68*(6), 1608–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7462458

Dreschler, W. A., & Plomp, R. (1985). Relations between psychophysical data and speech perception for hearing-impaired subjects. II. *The Journal of the Acoustical Society of America*, *78*(4), 1261–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4056220

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of Acoustical Society of America*, *76*(1), 87–96.

Dupont, W. D., & Plummer, W. D. (1990). Power and sample size calculations: A review and computer program. *Controlled Clinical Trials*, *11*, 116–128.

Durlach, N. I., Mason, C. R., Kidd, G. R., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America*, *113*(6), 2984. doi:10.1121/1.1570435

Evans, E. F. (1967). Narrow “tuning” of cochlear nerve fibre responses in the guinea-pig. *Physiology*, *206*, 14–15P.

Evans, E. F. (1975). The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea. *Audiology□: Official Organ of the International Society of Audiology*. doi:10.3109/00206097509071754

Fletcher, H. (1950). A Method of Calculating Hearing Loss for Speech from an Audiogram. *Journal of Acoustical Society of America*, *22*(1), 1–5.

Fletcher, H., & Steinberg, J. C. (1947). Articulation testing methods. *Bell Telephone Laboratories*, 1–98.

Fontan, L., Tardieu, J., Gailaird, P., Woisard, V., & Ruiz, R. (2015). Relationship Between Speech Intelligibility and Speech Comprehension in Babble Noise. *Journal of Speech, Language and Hearing Research*, *58*(June), 977–986. doi:10.1044/2015

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*, 113–133.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct- realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongue. *Journal of Acoustical Society of America*, *99*(3), 1730–1741.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5), 2246. doi:10.1121/1.1689343

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Fu, C., & Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hearing Research*, *211*, 74–84. doi:10.1016/j.heares.2005.09.001

Glasberg, B. R., & Moore, B. C. . (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology. Supplementum*, *32*(November 2015), 1–25.

Glasberg, B. R., & Moore, B. C. . (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*, 103–138.

Golding, M., Seymour, J., Dillon, H., & Carter, L. (2007). *The development of a telephone-based screener of hearing disability*.

Grant, K. W. (2002). Measures of auditory-visual integration for speech understanding: A theoretical perspective (L). *The Journal of the Acoustical Society of America*, *112*(1), 30. doi:10.1121/1.1482076

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, *103*(5 Pt 1), 2677–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9604361

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise.

Scandinavian Audiology, (11), 1982.

Hall, S. J. (2006). *The development of a new English sentence in noise test and an English number recognition test.* Univeristy of Southampton.

Harris, F. P. (1990). Distortion-product otoacoustic emissions in human with high frequency sensorineural hearing loss. *Journal of Speech , Language and Hearing Research*, *33*, 594–600.

HearCom. (2005). *Hearing in the Communication Society D-1-2□: Report on the proposed set of communication performance tests*.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *The Journal of Speech and Hearing Disorders*, *17*(3), 321–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13053556

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology*, *51*(7), 536–44. doi:10.3109/14992027.2012.670731

Hochmuth, S., Jürgens, T., Brand, T., & Kollmeier, B. (2015). Talker- and language-specific effects on speech intelligibility in noise assessed with bilingual talkers: Which language is more robust against noise and reverberation? *International Journal of Audiology*, *2027*(October), 1–12. doi:10.3109/14992027.2015.1088174

Hochmuth, S., Kollmeier, B., Brand, T., & Jürgens, T. (2015). Influence of noise type on speech reception thresholds across four languages measured with matrix sentence tests. *International Journal of Audiology*, (April), 1–9. doi:10.3109/14992027.2015.1046502

Hong, Y. P. (1984). Disyllabic Malay word lists for speech audiometry. *Medical Journal of Malaysia*, *39*(3), 197–204.

Hope, R. V. (2010). *Towards the Development of the New Zealand Heairng in Noise test (NZHINT).* University of Canterbury, New Zealand.

Houben, R., & Dreschler, W. A. (2015). Optimization of the dutch matrix test by random selection of sentences from a preselected subset. *Trends in Hearing*, *19*, 1–10. doi:10.1177/2331216515583138

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., Verschuure, H., & Dreschler, W. A. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. *International Journal of Audiology*, *53*(10), 760–3. doi:10.3109/14992027.2014.920111

Humes, L. E., Dirks, D. D., Bell, T. S., & Kincaid, G. E. (1987). Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. *The Journal of the Acoustical Society of America*, *81*(3), 765–773. doi:10.1121/1.394845

Huxley, A. (1990). Bone conducted sounds. *Nature*, *343*, 28`.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of Acoustical Society of America*, *97*(1), 553–562. doi:10.1121/1.412280

Jang, Y., Mortimer, J. A., Haley, W. E., Chisolm, T. E. H., & Graves, A. B. (2002). Nonauditory Determinants of Self-Perceived Hearing Problems Among Older Adults□: The Role of Stressful Life Conditions , Neuroticism , and Social Resources. *Journal of Gerontology*, *57*(7), 466–469.

Jansen, S., Luts, H., Wagener, K. C., Frachet, B., & Wouters, J. (2010). The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *International Journal of Audiology*, *49*(5), 378–87. doi:10.3109/14992020903431272

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., … van Wieringen, A. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *International Journal of Audiology*, *51*(3), 164–173. doi:10.3109/14992027.2011.633568

Jones, J. a, & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129–33. doi:10.1097/01.wnr.0000074343.81633.2a

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351.

Kendall, D. (1953). Audiometry for young children part 1. *Teacher of the Deaf*, *51*, 171–178.

Killion, M. C. (1997). The SIN report: Circuits haven’t solved the hearing in noise problem. *The Hearing Journal*, *50*(10), 28–32.

Killion, M. C., & Niquette, P. A. (2000). What can pure-tone audiogram tell about a patient’s SNR loss? *The Hearing Journal*, *53*(3), 46–53.

King, S. M. (2010). *Development and evaluation of a New Zealand digit triplet test for auditory screening*. University of Canterbury.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, *early onli*(September), 1–14. doi:10.3109/14992027.2015.1020971

Kramer, S. E., Kapteyn, T. S., & Festen, J. M. (1998). The self-reported Handicaping effect of hearing dissabilities. *Audiology*, *37*(5), 302–312.

Kuhl, P. K., & Iverson, P. (1995). Linguistic Experience and the “Perceptual Magnet Effect.” In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (2nd ed., pp. 121–154). Maryland: York Press.

Kujala, T., Tiitinen, H., Alku, P., Lehtokoski, A., & Ilmniemi, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 638–642.

Ladefoged, P., & Johnson, K. (2014). *A Course in Phonetics*. Cengage learning.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920–33. doi:10.1038/nrn2532

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi:10.1177/026553229901600103

Leensen, M. C. J., de Laat, J. a P. M., & Dreschler, W. A. (2011). Speech-in-noise screening tests by internet, part 1: test evaluation for noise-induced hearing loss identification. *International Journal of Audiology*, *50*(11), 823–34. doi:10.3109/14992027.2011.595016

Leensen, M. C. J., de Laat, J. a P. M., Snik, A. F. M., & Dreschler, W. A. (2011). Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss. *International Journal of Audiology*, *50*(11), 835–48. doi:10.3109/14992027.2011.595017

Leensen, M. C. J., & Dreschler, W. A. (2013). Speech-in-noise screening tests by internet, part 3: test sensitivity for uncontrolled parameters in domestic usage. *International Journal of Audiology*, *52*(10), 658–69. doi:10.3109/14992027.2013.803610

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), Suppl 2:467+. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5541744

Levitt, H., & Resnick, S. (1978). Speech reception by the hearing-impaired: methods of testing and the development of new tests. *Scandinavian Audiology*, *6*, 107–130.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *The Journal of the Acoustical Society of America*, *74*(6), 43–461.

Liberman, M. C., & Guinan, J. J. (1998). Feedback control of the auditory periphery: anti masking effects of the middle ear muscles and the olivocochlear efferents. *Journal of Communication Disorders*, *31*(617), 471–483.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics* (1st ed.). Cambridge University Press.

Linssen, A. M., Anteunis, L. J. C., & Joore, M. a. (2015). The Cost-Effectiveness of Different Hearing Screening Strategies for 50- to 70-Year-Old Adults: A Markov Model. *Value in Health*, 1–10. doi:10.1016/j.jval.2015.03.1789

Lutfi, R. A. (1990). How much masking is informational masking? *The Journal of the Acoustical Society of America*, *88*(6), 2607–2610. doi:10.1121/1.399980

Macdonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, *24*(3), 253–257.

MacPherson, A., & Akeroyd, M. A. (2014). Variations in the Slope of the Psychometric Functions for Speech Intelligibility: A Systematic Survey. *Trends in Amplification*, *18*, 10–16. doi:10.1177/2331216514537722

Massaro, D. W., Cohen, G., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America*, *67*(3), 996. doi:10.1121/1.383941

McClelland, A. D. (2015). *Refinement and Normalisation of the University of Canterbury Auditory-Visual Matrix Sentence Test*. University of Canterbury.

Mcclelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(23/30), 746–748.

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology*, *47*(9), 546–53. doi:10.1080/14992020802252261

Meyer, C., & Hickson, L. (2012). What factors influence help-seeking for hearing impairment and hearing aid adoption in older adults? *International Journal of Audiology*, *51*(2), 66–74. doi:10.3109/14992027.2011.611178

Meyer, C., Hickson, L., Khan, A., Hartley, D., Dillon, H., & Seymour, J. (2011). Investigation of the Actions Taken by Adults Who Failed a Telephone-Based Hearing Screen. *Ear and Hearing*, *32*, 720–731.

Miller, B. Y. G. A., Heise, G. A., & Lighten, W. (1950). The intelligibility of speech as a function of the test materials. *Harvard University*, (April), 329–335.

Moore, B. C. . (1996). Perceptual Consequences of Cochlear Hearing Loss and their Implications for the Design of Hearing Aids. *Ear and Hearing*, 133:161.

Moore, B. C. . (2001). Dead Regions in the Cochlea: Diagnosis, Perceptual Consequences, and Implications for the Fitting of Hearing Aids. *Trends in Amplification*, *5*(1), 1–34. doi:10.1177/108471380100500102

Moore, B. C. . (2003). *An Introduction to the Psychology of Hearing*. *Boston Academic Press* (Vol. 3). doi:10.1016/j.tins.2007.05.005

Moore, B. C. ., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, *188*(1-2), 70–88. doi:10.1016/S0378-5955(03)00347-2

Moore, B. C. ., Huss, M., Vickers, D. A., Glasberg, B. R., & Alcantara, J. I. (2000). A Test for the Diagnosis of Dead Regions in the Cochlea. *British Journal of Audiology*, *34*(4), 205–224.

Moulin, A., Bera, J. C., & Collet, L. (1994). Distortion product otoacoustic emissions and sensorineural hearing loss. *Audiology*, *33*(November), 305–326. doi:10.3109/00206099409071890

Mukari, S. Z. S., Haniza, N., Wahat, A., & Mazlan, R. (2014). Effects of Ageing and

Hearing Thresholds on Speech Perception in Quiet and in Noise Perceived in Different Locations. *Korean Journal of Audiology*, *18*(3), 112–118. doi:10.7874/kja.2014.18.3.112

Mukari, S. Z.-M. S., Keith, R. W., Tharpe, A. M., & Johnson, C. D. (2006). Development and standardization of single and double dichotic digit tests in the Malay language. *International Journal of Audiology*, *45*(6), 344–52. doi:10.1080/14992020600582174

Mukari, S. Z.-M. S., & Said, H. (1991). The development of Malay speech audiometry. *Medical Journal of Malaysia*, *46*(3), 262–268.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–78. doi:10.1121/1.397688

Murray, C. (2012). *Development of a Maori Language Version of the New Zealand Hearing Screening Test*. University of Canterbury.

Näätänen, R., & Kreegipuu, K. (2012). The mismatch negativity. In S. J. Luck & E. S. Kappenmen (Eds.), *The Oxford Handbook of Event Related Potential Components* (pp. 143–157). Oxford, New York: Oxford University Press.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtilainen, M., Livonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Nachtegaal, J., Kuik, D. J., Anema, J. R., Go, S. T., Festen, J. M., & Kramer, S. E. (2009). Hearing status , need for recovery after work , and psychosocial work characteristics□: Results from an internet-based national survey on hearing. *International Journal of Audiology*, *48*, 684–691. doi:10.1080/14992020902962421

Najibah Mahirah Awang. (2015). *Preliminary investigation of standard Malay vowels, first and second formants.* International Islamic University Malaysia.

Neely, S. T., & Kim, D. O. (1983). An active cochlear model showing sharp tuning and high sensitivity. *Hearing Research*, *9*(2), 123–130. doi:10.1016/0378-5955(83)90022-9

Neely, S. T., & Kim, D. O. (1986). Model for active elements in cochlear biomechanics. *Journal of Acoustical Society of America*, *79*(5), 1472–1480.

Nilsson, M., Soli, S. D., & Sullivan, J. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise The BKB sentence materials. *The Journal of the Acoustical Society of America*, *95*(June 1993), 1085–1099.

O'Beirne, G., King, S. M., Murray, C., Fromont, R., & King, J. (2012). Development and evaluation of the New Zealand Hearing Screening test. In *Audiological Society of Australia COnference*. Adelaide, Australia.

O'Beirne, G., Trounson, R. T., McClelland, A. D., & Maclagan, M. A. (2015). Development of an auditory-visual matrix sentence test in New Zealand English.

In *12th European Federation of Audiological Societies Congress*. Istanbul.

Osterhout, L., & Holcomb, P. J. (1993). Event-related Potentials and Syntactic Anomaly□: Evidence of Anomaly Detection During the Perception of Continuous Speech. *Language and Cognitive Processes*, *8*(4), 413–437.

Ozimek, E., Kutzner, D., & Libiszewski, P. (2012). Speech intelligibility tested by the Pediatric Matrix Sentence test in 3–6year old children. *Speech Communication*, *54*(10), 1121–1131. doi:10.1016/j.specom.2012.06.001

Ozimek, E., Kutzner, D., Sek, A., & Wicher, A. (2008). Accepted Manuscript. *Speech Communication*, *4*(2009).

Ozimek, E., Kutzner, D., Sek, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology*, *48*(7), 433–43. doi:10.1080/14992020902725521

Ozimek, E., Kutzner, D., Sęk, A., & Wicher, A. (2009). Development and evaluation of Polish digit triplet test for auditory screening. *Speech Communication*, *51*(4), 307–316. doi:10.1016/j.specom.2008.09.007

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology*, *49*(6), 444–454. doi:10.3109/14992021003681030

Ozog, a. C. K. (1987). The Syntax of the Mixed Language of Malay -- English Bilinguals. *RELC Journal*, *18*(1), 72–90. doi:10.1177/003368828701800105

Patuzzi, R. B., Yates, G. K., & Johnstone, B. M. (1989). Outer hair cell receptor current and sensorineural hearing loss. *Hear Res*, *42*(1), 47–72. doi:0378-5955(89)90117-2 [pii]

Peters, R. W., Moore, B. C. ., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of Acoustical Society of America*, *103*(1), 577–587.

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, *42*(2), 26–32.

Plant, G., Phillips, D. P., & Tsembis, J. (1982). An auditory-visual speech test for the elderly hearing impaired. *Australian Journal of Audiology*, *4*, 62–68.

Plomp, R. (1994). Noise, amplification, and compression: considerations of three main issues in hearing aid design. *Ear and Hearing*, *15*(1), 2–12. doi:10.1097/00003446-199402000-00002

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, *18*, 43–52.

Pollack, I. (1975). Auditory informational masking. *Journal of Acoustical Society of America*, *57*(supp 1), 5.

Pronk, M., Deeg, D. J. H., Festen, J. M., Twisk, J. W., Smits, C., Comijs, H. C., & Kramer, S. E. (2013). Decline in Older Persons ' Ability to Recognize Speech in Noise□: The Influence of Demographic , Health-Related , Environmental , and

Cognitive Factors. *Ear and Hearing*, *34*(6), 722–732.

Puglisi, G. E., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., & Kollmeier, B. (2015). An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *International Journal of Audiology*, *early onli*(September), 1–7. doi:10.3109/14992027.2015.1061709

Quar, T. K., Mukari, S. Z.-M. S., Abdul Wahab, N. A., Abdul Razak, R., Omar, M., & Maamor, N. (2008). The Malay hearing in noise test. *International Journal of Audiology*, *47*(6), 379–80. doi:10.1080/14992020801886796

Rance, G., Corben, L., Barker, E., Carew, P., Chisari, D., Rogers, M., … Delatycki, M. B. (2010). Auditory perception in individuals with friedreich's ataxia. *Audiology and Neurotology*, *15*, 229–240. doi:10.1159/000255341

Raof, W. R. W. A. (2010). *Population and Basic Demographic Characteristic Report 2010*. *Census 2010*. Retrieved from https://www.statistics.gov.my/index.php?r=column/ctheme&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09&bul_id=MDMxdHZjWTk1SjFzTzNkRXYzcVZjdz09

Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, *80*(1), 4–9. doi:10.1016/j.biopsycho.2008.05.002

Rönnberg, J., Hygge, S., Keidser, G., & Rudner, M. (2014). The effect of functional hearing loss and age on long- and short-term visuospatial memory: evidence from the UK biobank resource. *Frontiers in Aging Neuroscience*, *6*(December), 326. doi:10.3389/fnagi.2014.00326

Rosdina, A., Leelavathi, M., Zaitun, A., Lee, V., Aizimah, M., Majmin, S., & Mohd, K. (2010). Self Reported Hearing Loss Among Elderly Malaysians. *Malaysian Family Medicine*, *5*(2), 91–94.

Royal National Institute for Deaf People. (2015). Action on hearing loss. Retrieved November 18, 2015, from http://www.actiononhearingloss.org.uk/about-us.aspx

Schädler, M. R., Warzybok, A., Hochmuth, S., & Kollmeier, B. (2015). Matrix sentence intelligibility prediction using an automatic speech recognition system. *International Journal of Audiology*, *early onli*(September), 1–8. doi:10.3109/14992027.2015.1061708

Schell-Majoor, L., Rennies, J., Ewert, S. D., & Kollmeier, B. (2015). Application of psychophysical models for audibility prediction of technical signals in real-world background noise. *Applied Acoustics*, *88*, 44–51. doi:10.1016/j.apacoust.2014.08.001

Shaw, E. A. G. (1968). Sound pressure generated in an External-Ear replica and real human ears by a nearby point source. *Journal of Acoustical Society of America*, *44*(1), 240–249.

Sincock, B. P. (2008). *Clinical applicability of adaptive speech testing: A comparison* *of the administration time, accuracy, efficiency and reliability of adaptive speech tests with conventional speech audiometry*. University of Canterbury.

Smits, C., Goverts, S. T., & Festen, J. M. (2013). The digits-in-noise test□: Assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America*, *133*(March), 1693–1706.

Smits, C., & Houtgast, T. (2005). Results From the Dutch Speech-in-Noise Screening Test by Telephone. *Ear and Hearing*, *26*, 89–95.

Smits, C., & Houtgast, T. (2006). Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests. *The Journal of the Acoustical Society of America*, *120*(3), 1608. doi:10.1121/1.2221405

Smits, C., & Houtgast, T. (2007). Recognition of digits in different types of noise by normal-hearing and hearing-impaired listeners. *International Journal of Audiology*, *46*(3), 134–144. doi:10.1080/14992020601102170

Smits, C., Kapteyn, T. S., & Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *International Journal of Audiology*, *43*, 15–28.

Smits, C., Kramer, S. E., & Houtgast, T. (2006). Speech reception thresholds in noise and self-reported hearing disability in a general adult population. *Ear and Hearing*, *27*(5), 538–49. doi:10.1097/01.aud.0000233917.72551.cf

Smits, C., Merkus, P., & Houtgast, T. (2006). How we do it□: The Dutch functional hearing – screening tests by telephone and internet. *Clinical Otolaryngology*, *31*, 436–440.

Soli, S. D., & Wong, L. (2008). Assessment of speech intelligibility in noise with the Hearing in Noise Test. *International Journal of Audiology*, *47*(December), 356–361. doi:10.1080/14992020801895136

Stam, M., Kostense, P. J., Festen, J. M., & Kramer, S. E. (2013). The relationship between hearing status and the participation in different categories of work: demographics. *Work (Reading, Mass.)*, *46*(2), 207–19. doi:10.3233/WOR-131747

Stam, M., Smits, C., Twisk, J. W., Lemke, U., Festen, J. M., & Kramer, S. E. (2015). Deterioration of Speech Recognition Ability Over a Period of 5 Years in Adults Aged 18 to 70 Years□: Results of the Dutch Online Speech-in-Noise Test. *Ear and Hearing*, *in press*, 1–9.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–62. doi:10.1016/j.bandl.2011.07.001

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45. doi:10.1109/ICSLP.1996.607202

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*(3), 234–249.

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of Acoustical Society of America*, *26*(2), 212–215.

Summerfield, A. Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, *335*(1273), 71–78.

Suter, A. H. (1985). Speech recognition in noise by individuals with mild hearing impairments. *The Journal of the Acoustical Society of America*, *78*(September), 887–900.

Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception — New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, *97*(1), 23–29. doi:10.1016/j.ijpsycho.2015.04.020

Taylor, B. B. (2003). Speech-in-noise tests□: How and why to include them in your basic test battery. *The Hearing Journal*, *56*(1), 40–46.

Teranishi, R., & Shaw, E. A. G. (1968). External-Ear Acoustic Models with Simple Geometry. *Journal of Acoustical Society of America*, *44*(1), 257–263.

Theelen-van den Hoek, F. L., Houben, R., & Dreschler, W. A. (2014). Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users. *International Journal of Audiology*, *53*(11), 817–28. doi:10.3109/14992027.2014.922223

Trounson, R. T., & O'Beirne, G. (2012). *Development of the UC auditory visual matrix sentence test*. University of Canterbury.

Tye-Murray, N. (1992). Preparing for Communication Interactions□: The Value of Anticipatory Strategies for Adults With Hearing Impairment. *Journal of Speech and Hearing Research*, *35*, 430–435.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007a). Audiovisual Integration and Lipreading Abilities of Older Adults with Normal and Impaired Hearing. *Ear and Hearing*, *28*(5), 656–668.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007b). Auditory and Visual Lexical Neighborhoods in Audiovisual Speech Perception. *Trends in Amplification*, *11*(4), 233–242.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007c). The Effects of Age and Gender on Lipreading Abilities. *Journal of American Academy of Audiology*, *892*, 883–892.

Tye-Murray, N., Sommers, M. S., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology*, *47 Suppl 2*, S31–7. doi:10.1080/14992020802301662

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., & Giguère, C. (2008). The Canadian French hearing in noise test. *International Journal of Audiology*, *47*(6), 383–5. doi:10.1080/14992020802055300

Van Engen, K. J., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing speech intelligibility: interactions among context, modality, speech style, and masker. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1908–18. doi:10.1044/JSLHR-H-13-0076

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of Acoustical Society of America*, *107*(3), 1671–1684. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10738820

Vestergaard Knudsen, L., Oberg, M., Nielsen, C., Naylor, G., & Kramer, S. E. (2010). Factors Influencing Help Seeking, Hearing Aid Uptake, Hearing Aid Use and Satisfaction With Hearing Aids: A Review of the Literature. *Trends in Amplification*, *14*(3), 127–154. doi:10.1177/1084713810385712

Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, *44*(3), 144–156. doi:10.1080/14992020500057517

Wagener, K. C., Brand, T., & Kollmeier, B. (1999). Entwicklung eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburg Satztests. *Z Audiol*, *38*(3), 86–95.

Wagener, K. C., Eeenboom, F., Brand, T., & Kollmeier, B. (2005). Ziffern-Tripel-Test: Sprachverständlichkeitstest über das Telefon. *DGA Jahrestagung*, 1–3.

Wagener, K. C., Josvassen, J. L., & Ardenkjaer, R. (2003). Design, optimization and evaluation of a Danish sentence in noise. *International Journal of Audiology*, *42*, 10–17.

Wagener, K. C., Kuhnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. *Z Audiol*, *38*(2), 5–15.

Wagener, K. C., Kuhnel, V., Kollmeier, B., Brand, T., Kollmeier, B., Kuhnel, V., … Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil II: Optimierung des Oldenburger Satztests Development and evaluation of a German sentence test Part II: Optimization of the Oldenburg sentence test. *Z Audiol*, *38*(2), 44–56.

Walden, B. E., Schwartz, D. M., Williams, D. L., Holum-hardegen, L. L., & Crowley, J. M. (1983). Test of the assumptions underlying comparative hearing aid evaluations. *Journal of Speech and Hearing Disorders*, *48*(August), 264–273.

Wan Aslynn, W. A. (2012). *Malay and Its Morphophonology in Vowel length discrimination among Malay speakers of Malaysian English: an instrumental study*. University of Newcastle Upontyne.

Wardenga, N., Batsoulis, C., Wagener, K. C., Brand, T., Lenarz, T., & Maier, H. (2015). Do you hear the noise? The German matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment.

International Journal of Audiology, *2027*(November), 1–9. doi:10.3109/14992027.2015.1079929

Warzybok, A., Brand, T., Wagener, K. C., & Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *International Journal of Audiology*, (early online), 1–12. doi:10.3109/14992027.2015.1063715

Warzybok, A., Hochmuth, S., Laan, M. Z. Der, & Kollmeier, B. (2005). New speech intelligibility tests for Spanish and Russian.

Warzybok, A., Zokoll, M. A., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test. *International Journal of Audiology*, 1–9. doi:10.3109/14992027.2015.1020969

Watson, C. S. (2005). Masking on Informational Some Comments. *Acta Acustica United with Acustica*, *91*(2005), 502–512.

Watson, C. S., Kidd, G. R., Miller, J. D., Smits, C., & Humes, L. E. (2013). Telephone screening tests for functionally impaired hearing: current use in seven countries and development of a US version. *Journal of the American Academy of Audiology*, *23*(10), 757–67. doi:10.3766/jaaa.23.10.2

WHO. (2012). *WHO global estimates on prevalence of hearing loss Mortality and Burden of Diseases*. Retrieved from http://www.who.int/pbd/deafness/WHO_GE_HL.pdf

WHO, S. A. (2007). *Situation Review and Update on Deafness , Hearing Loss and Intervention Programmes Proposed Plans of Action for Prevention*.

Williams-Sanchez, V., Mcardle, R. a, Wilson, R. H., Kidd, G. R., Watson, C. S., & Bourne, A. L. (2014). Validation of a Screening Test of Auditory Function Using the Telephone. *Journal of the American Academy of Audiology*, *25*(10), 937–951. doi:10.3766/jaaa.25.10.3

Wilson, R. H., Burks, C. A., & Weakley, D. G. (2005). A comparison of word-recognition abilities assessed with digit pairs and digit triplets in multitalker babble. *The Journal of Rehabilitation Research and Development*, *42*(4), 499. doi:10.1682/JRRD.2004.10.0134

Wolfe, J. (n.d.). Voice Acoustics: an introduction.

Wong, L., Liu, S., & Han, N. (2008). The Mainland Mandarin hearing in noise test. *International Journal of Audiology*, *47*(6), 393–395.

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: clinical implications. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *44*(3), 253–70. doi:10.1080/13682820802090281

World Bank Group. (2014). Internet users (per 100 people).

Yueh, B., Collins, Ã. M. P., Souza, P. E., Boyko, E. J., Loovis, C. F., & Heagerty, P. J. (2010). Long-Term Effectiveness of Screening for Hearing Loss: The

Screening for Auditory Impairment F Which Hearing Assessment Test (SAI-WHAT) Randomized Trial. *Journal of American Geriatric Society*, *58*, 427–434. doi:10.1111/j.1532-5415.2010.02738.x

Yunus, M. (1980). *The Malay sound system* (2nd Editio.). Kuala Lumpur: Fajar Bakti.

Zera, J. (2004). Speech intelligibility measured by adaptive maximum-likelihood procedure. *Speech Communication*, *42*(3-4), 313–328. doi:10.1016/j.specom.2003.08.007

Zokoll, M. A., Fidan, D., Türkyılmaz, D., Hochmuth, S., Ergenç, İ., Sennaroğlu, G., & Kollmeier, B. (2015). Development and evaluation of the Turkish matrix sentence test. *International Journal of Audiology*, *2027*(October), 1–11. doi:10.3109/14992027.2015.1074735

Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-Noise Tests for Multilingual Hearing Screening and Diagnostics. *American Journal of Audiology*, *22*(June), 175–178. doi:10.1044/1059-0889(2013/12-0061)American

Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M., Kollmeier, B., & Buscherm, M. (2012). Internationally comparable screening tests for listening in noise in several European languages□: The German digit triplet test as an optimization prototype Key Words□: *International Journal of Audiology*, *51*(January), 697–707. doi:10.3109/14992027.2012.690078

Zurek, P. M., & Delhorne, L. A. (1987). Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *Journal of Acoustical Society of America*, *82*(5), 1548–1559.

APPENDIX A: List of 100 sentences containing all possible word combinations

Saya	bagi	satu	bola	baru
Kita	bagi	dua	bola	besar
Dia	bagi	banyak	bola	lama
Kami	bagi	semua	bola	kecil
Ibu	bagi	tiga	bola	merah
Abang	bagi	empat	bola	hitam
Ayah	bagi	lima	bola	putih
Adik	bagi	enam	bola	hijau
Kakak	bagi	tujuh	bola	mahal
Nenek	bagi	lapan	bola	cantik
Saya	ada	satu	buku	baru
Kita	ada	dua	buku	besar
Dia	ada	banyak	buku	lama
Kami	ada	semua	buku	kecil
Ibu	ada	tiga	buku	merah
Abang	ada	empat	buku	hitam
Ayah	ada	lima	buku	putih
Adik	ada	enam	buku	hijau
Kakak	ada	tujuh	buku	mahal
Nenek	ada	lapan	buku	cantik
Saya	dapat	satu	baju	baru

Kita	dapat	dua	baju	besar
Dia	dapat	banyak	baju	lama
Kami	dapat	semua	baju	kecil
Ibu	dapat	tiga	baju	merah
Abang	dapat	empat	baju	hitam
Ayah	dapat	lima	baju	putih
Adik	dapat	enam	baju	hijau
Kakak	dapat	tujuh	baju	mahal
Nenek	dapat	lapan	baju	cantik
Saya	perlu	satu	lampu	baru
Kita	perlu	dua	lampu	besar
Dia	perlu	banyak	lampu	lama
Kami	perlu	semua	lampu	kecil
Ibu	perlu	tiga	lampu	merah
Abang	perlu	empat	lampu	hitam
Ayah	perlu	lima	lampu	putih
Adik	perlu	enam	lampu	hijau
Kakak	perlu	tujuh	lampu	mahal
Nenek	perlu	lapan	lampu	cantik
Saya	beri	satu	meja	baru
Kita	beri	dua	meja	besar
Dia	beri	banyak	meja	lama
Kami	beri	semua	meja	kecil
Ibu	beri	tiga	meja	merah

Abang	beri	empat	meja	hitam
Ayah	beri	lima	meja	putih
Adik	beri	enam	meja	hijau
Kakak	beri	tujuh	meja	mahal
Nenek	beri	lapan	meja	cantik
Saya	ambil	satu	kotak	baru
Kita	ambil	dua	kotak	besar
Dia	ambil	banyak	kotak	lama
Kami	ambil	semua	kotak	kecil
Ibu	ambil	tiga	kotak	merah
Abang	ambil	empat	kotak	hitam
Ayah	ambil	lima	kotak	putih
Adik	ambil	enam	kotak	hijau
Kakak	ambil	tujuh	kotak	mahal
Nenek	ambil	lapan	kotak	cantik
Saya	mahu	satu	kunci	baru
Kita	mahu	dua	kunci	besar
Dia	mahu	banyak	kunci	lama
Kami	mahu	semua	kunci	kecil
Ibu	mahu	tiga	kunci	merah
Abang	mahu	empat	kunci	hitam
Ayah	mahu	lima	kunci	putih
Adik	mahu	enam	kunci	hijau
Kakak	mahu	tujuh	kunci	mahal

Nenek	mahu	lapan	kunci	cantik
Saya	suka	satu	pisau	baru
Kita	suka	dua	pisau	besar
Dia	suka	banyak	pisau	lama
Kami	suka	semua	pisau	kecil
Ibu	suka	tiga	pisau	merah
Abang	suka	empat	pisau	hitam
Ayah	suka	lima	pisau	putih
Adik	suka	enam	pisau	hijau
Kakak	suka	tujuh	pisau	mahal
Nenek	suka	lapan	pisau	cantik
Saya	nampak	satu	mangkuk	baru
Kita	nampak	dua	mangkuk	besar
Dia	nampak	banyak	mangkuk	lama
Kami	nampak	semua	mangkuk	kecil
Ibu	nampak	tiga	mangkuk	merah
Abang	nampak	empat	mangkuk	hitam
Ayah	nampak	lima	mangkuk	putih
Adik	nampak	enam	mangkuk	hijau
Kakak	nampak	tujuh	mangkuk	mahal
Nenek	nampak	lapan	mangkuk	cantik
Saya	minta	satu	topi	baru
Kita	minta	dua	topi	besar
Dia	minta	banyak	topi	lama

Kami	minta	semua	topi	kecil
Ibu	minta	tiga	topi	merah
Abang	minta	empat	topi	hitam
Ayah	minta	lima	topi	putih
Adik	minta	enam	topi	hijau
Kakak	minta	tujuh	topi	mahal
Nenek	minta	lapan	topi	cantik

APPENDIX B: Phonetic transcription of top 1000 words in the Malay corpus.

1000 most frequent words used in Malay daily	Consonant																							Vowels						Diphthongs			SUM
	Bilabial			Labio-dental		Alveolar							Post-alveolar			Palatal		Velar			Glottal		Other	Front		Mid		Back					
	Nasal	Plosive		Fricative		Nasal	Plosive		Fricative		Approximant	Lateral	Affricate		Fricative	Nasal	Approximant	Nasal	Plosive		Fricative	Stop											
Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	
yang																	1	1									1						3
dan						1		1																			1						3
di								1																1									2
itu							1																	1				1					3
dengan						1		1										1							1		1						5

ini						1																		2									3

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
dalam	1							1				1															2						5
pada		1						1																	1		1						4
untuk						1	1															1						1	1				5
beliau			1									1												1	1		1	1					6
oleh												1									1					1			1				4
kita							1												1					1	1								4
tidak							1	1														1		1			1						5
telah							1					1									1				1		1						5
katanya							1									1			1						2		1						6
daripada		1						2			1													1	1		2						8
ke																			1						1								2
atau							1																				1				1		3

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
saya									1								1								1		1						4
akan						1													1								2						4
boleh			1									1									1					1			1				5
mereka	1										1								1						2	1							6
juga														1						1					1			1					4
ada								1																	1		1						3
tahun						1	1														1						1	1					5
ialah												1									1			1	1		1						5
kepada		1						1											1						2		1						6
berkata			1				1				1								1						2		1						7
satu							1		1																		1	1					4
dari								1			1													1			1						4
negara						1					1									1					2		1						6
sebagai			1						1											1					1					1			5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
dapat		1						2																			2						5
menjadi	1					1		1						1										1	1		1						7
seperti		1					1		1															1	2								6
tersebut			1				1		1		1														1			1					6
lebih			1									1									1				1	1							5
tetapi		1					2																	1	1		1						6
adalah								1				1									1				1		3						7
lain						1						1												1			1						4
kata							1												1						1		1						4
kerajaan						1					1			1					1						1		3						8
kerana						1					1								1						2		1						6
ia																								1	1								2
menteri																																	0
negeri						1					1									1				1	2								6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
juta							1							1													1	1					4
baru			1								1																1	1					4
beberapa		1	2								1														3		1						8
pula		1										1													1			1					4
syarikat							1								1				1					1			2						6
apabila		1	1									1												1	2		1						7
jika														1					1					1	1								4
lalu												2															1	1					4
datuk							1	1														1					1		1				5
malaysia	1								1			1												1	2	1							7
perlu		1									1	1													1			1					5
bagi			1																	1				1			1						4
dua								1															1		1			1					4
lagi												1								1				1			1						4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
dia								1																1	1								3
banyak			1													1						1					2						5
antara						1	1				1														1		2						6
mempunyai	2	1														1								1	1		1	1					8
hari											1										1			1			1						4
kuala												1							1						1		1	1					5
antara						1	1				1														1		2						6
apabila		1	1									1												1	2		1						7
selepas		1							2			1													2		1						7
pihak		1																			1	1				1	1						5
sama	1								1																1		1						4
mengenai	1					1												1						1	1		1						6
lumpur	1	1										1																1	1				5
ekonomi	1					1																		1		1			2				6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
ketika							1												2					1	2								6
masa	1								1																1		1						4
doktor							1	1														1			1				1				5
secara									1		1		1												2		1						6
kini						1													1					2									4
turut							1	1			1																	2					5
masih	1								1												1					1	1						5
sebelum	1		1						1			1													2			1					7
bahawa			1																		1		1		1		2						6
seorang									1		1							1							1		1		1				6
memberi	2		1								1													1	2								7
hanaya																1					1				1		1						4
sudah								1	1												1						1	1					5
dunia						1		1																1	1			1					5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐ u	oɪ	Sum
kami	1																		1					1			1						4
para																																	0
semua	1								1																2			1					5
peratus		1					1		1		1														1		1	1					7
menurut	1					1		1			1														1			2					7
terhadap		1					1	1			1										1				1		2						8
masalah	1								1			1									1						4						8
bukan			1			1													1								1	1					5
sendiri						1		1	1		1													2	1								7
pemain	1	1				1																		1	1		1						6
raya											1						1								1		1						4
islam	1								1			1												1			1						5
rakyat								1			1						1					1					2						6
apa		1																							1		1						3
perdana		1				1		1			1														2		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
besar			1						1																1		1						4
sahaja									1					1							1				1		2						6
pelajar		1										1		1											1		2						6
sini						1			1															2									4
sebuah			1						1												1				1		1	1					6
merupakan	1	1				1													1						2		1	1					8
melayu	1											1					1								1		1	1					6
baik			1																			1			1		1						4
saham	1								1												1						2						5
luar												1															1	1					3
bagaimana pun	1	1	1			2														1					1		2	1		1			11
mata	1						1																		1		1						4
sukan						1			1										1								1	1					5
kumpulan	1	1				1						1							1								1	2					8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
pasukan		1				1			1										1								2	1					7
tinggi							1											1		1				2									5
iaitu							1										1							1			1	1					5
pertama	1	1					1				1														2		1						7
tiga							1													1				1	1								4
mungkin	1					1												1	1					1				1					6
membuat	2		1					1																	1		1	1					7
mengambi l	2		1									1						1						1	1		1						8
rumah	1										1										1						1	1					5
pasaran		1				1			1		1																3						7
walaupun		1				1						1											1				1	1			1		7
tempatan	1	1				1	2																		1		2						8
sebanyak			1						1							1						1			1		2						7
atas							1		1																		2						4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
majlis	1								1			1		1										1			1						6
selain						1			1			1													1					1			5
setiap		1					1		1															1	1		1						6
berkenaan			1			2					1								1						2		2						9
mendapat	1	1				1	1	1																	1		2						8
terus							1		1		1														1			1					5
mahu	1																				1						1	1					4
supaya		1							1								1								1		1	1					6
filem	1			1								1												1	1								5
masyarakat	1							1			1				1				1								4						9
melalui	1											2												1	1		1	1					7
keadaan						1		1											1						1		3						7
sekolah									1			1							1		1				1		1		1				7
diri								1			1													2									4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
berlaku			1								1	1							1						1		1	1					7
sehingga									1									1		1	1			1	2								7
wanita						1	1																1	1			2						6
seri									1		1													1	1								4
kewangan						1												1	1				1		1		2						7
termasuk	1						1		1		1											1			1		1		1				8
pelbagai		1	1									1								1					1		1			1			7
album	1		1									1															1	1					5
sejak									1					1								1			1		1						5
bidang			1					1										1						1			1						5
sementara	1					1	1		1		1														3		1						9
ramai	1										1																1			1			4
utama	1						1																		1		1	1					5
bulan			1			1						1															1	1					5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
usaha									1												1				2			1					5
sekarang									1		1							1	1						1		2						7
kali												1							1					1			1						4
polis		1							1			1												1					1				5
ketua							1												1						2			1					5
presiden		1				1		1	1		1													1	1	1							8
tempat	1	1					1	1																	1		1						6
pusat		1						1	1																		1	1					5
anak						1																1					2						4
wang																		1					1				1						3
semasa	1								2																2		1						6
kawasan						1			1										1				1				3						7
menerima	2					1					1													1	3								8
jumlah	1											1		1							1						1	1					6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
kedua								1											1						2			1					5
pernah		1				1					1										1				1		1						6
menggunakan	1					2												1	1	1					2		1	1					10
begitu			1				1													1				1	1			1					6
pendidikan		1				2		1											1			1		1	1	1	1						10
semalam	2								1			1													1		2						7
mohd	2							1													1						2	1					7
keputusan		1				1	1		1										1						1		1	2					9
pembangunan	1	1	1			2												1							1		2	1					10
ibu			1																					1				1					3
asia									1															1			2						4
antarabangsa			1			1	1		1		1							1							2		3						11
hingga																		1		1	1			1	1								5
tanpa	1	1					1																		1		1						5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ε**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
berbanding			2			1		1			1							1						1	1		1						9
perkara		1									2								1						2		1						7
indonesia						2		1	1															2		1	1		1				9
berjaya			1								1			1			1								2		1						7
semula	1								1			1													1			1					5
anwar						1																	1				2						4
sedang								1	1									1							1		1						5
depan		1				1		1																	1		1						5
sistem	1						1		2															1	1								6
program	1	1									2									1							1		1				7
peringkat		1						1			1							1	1					1	1		1						8
manakala	1					1						1							1						2		2						8
abdul			1					1				1															1	1					5
tindakan						1	1	1											1			1		1			2						8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
jalan						1						1		1													2						5
politik		1					1					1										1		2					1				7
malah	`											`									`						2						2
mampu	2	1																									2	1					6
bersama	1		1						1		1														2		1						7
bawah			1																		1		1				2						5
mahathir	1							1													1					1	2						6
namun	1					2																					1	1					5
parti		1					1				1													1			1						5
industri						1	1	1	1		1													2				1					8
pilihan		1				1						1									1			2			1						7
penting		1				1	1											1						1	1								6
lagu												1								1							1	1					4
bank			1															1	1							1							4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
perniagaan		1				2					1									1				1	1		3						10
membantu	2		1			1	1																		1		1	1					8
terpaksa		1					1		1		1											1			2		1						8
tentang						1	1											1							1		1						5
kementerian	1					2	1				1								1					1	1		1						9
selama	1								1			1													2		1						6
projek		1									1			1								1				1			1				6
mana	1					1																			1		1						4
harus									1		1										1						1	1					5
teknologi						1	1					1		1								1		1	1				2				9
anggota							1											1		1					1		1		1				6
kejohanan						2								1					1		1				1		2		1				9
sekali									1			1							1					1			1						5
lama	1											1													1		1						4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
universiti					1	1	1		1		1													3	1			1					10
empat	1	1					1																		1		1						5
setelah							1		1			1									1				2		1						7
bhd			1					1			1										1				1		1						6
berada			1					1			1														2		1						6
demikian	1					1		1											1					2	1		1						8
amat	1							1																			2						4
terdapat		1					1	2			1														1		2						8
umno	1					1																					1		1				4
pegawai		1																		1			1		1		1			1			6
mengadakan	1					1		1										1	1						2		2						9
memberitahu	2		1				1				1										1			1	2		1	1					11
harga											1									1	1				1		1						5
dilakukan						1		1				1							2					1			2	1					9

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
asing									1									1						1			1						4
kalau												1							1								1				1		4
belum	1		1									1													1			1					5
amerika	1										1								1					1	1	1	1						7
peluang		1										1						1							1		1	1					6
membawa	2		1																				1		2		1						7
timbalan	1		1			1	1					1												1			2						8
menghada pi	1	1						1										1			1			1	1		2						9
pengarah		1									1							1			1				1		2						7
jabatan			1			1	1							1													3						7
maklumat	2							1				1										1					2	1					8
pun		1				1																							1				3
tempoh	1	1					1														1				1				1				6
langkah												1						1	1		1						2						6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
akhir																					1					1	1						3
menarik	1					1					1											1		1	1		1						7
perkhidmatan	1	1				1	1	1			1										1			1	1		2						11
air																	1									1	1						3
awal												1											1				2						4
dibuat			1					2																1			1	1					6
persatuan		1				1	1		1		1														1		2	1					9
salah									1			1									1						2						5
kerja											1			1					1						2								5
akhbar			1																		1						2						4
ahmad	1							1													1						2						5
memberikan	2		1			1					1								1					1	2		1						10
singapura		1							1		1							1						1	2			1					8
bahasa			1						1												1				1		2						6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
kenyataan						1	1									1			1						1		3						8
minggu	1																	1		1				1				1					5
ahli												1									1			1			1						4
lima	1											1												1	1								4
malam	2											1															2						5
dijangka								1						1				1	1					1	1		1						7
pemimpin	2	2				1																		2	1								8
seluruh									1		1	1									1				1			1	1				7
tan						1	1																				1						3
bekas			1						1										1						1		1						5
mahkamah	2																		1		2						3						8
sebarang			1						1		1							1							1		2						7
penyanyi		1														2								1	1		1						6
semakin	1					1			1										1					1	1		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
paling		1										1						1						1			1						5
jelas									1			1		1											1		1						5
sebenarnya			1			1			1							1									3		1						8
terbaik			1				1				1											1		1	1		1						7
sesuatu							1		2																1		1	2					7
cara											1		1												1		1						4
diadakan						1		2											1					1	1		2						8
meningkat	1					1		1										1	1					1	1		1						8
bilion			1			1						1												2	1								6
kalangan						1						1						1	1								3						7
kes									1										1						1								3
latihan						1	1					1									1			1			2						7
mendapatkan	1	1				2		2											1						1		3						11
bernilai			1			1					1	1												1	1					1			7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
bahagian			1			1														1	1			1			3						8
keluarga											1	1							1	1			1		2		1	1					9
melihat	1							1				1									1			1	1		1						7
memastikan	2					1	1		1										1					1	1		2						10
kebangsaan			1			1			1							1			1						1		3						9
kadar								1			1								1								2						5
hati							1														1			1			1						4
isu									1															1				1					3
menerusi	1					1			1		1													1	2			1					8
keduadua								2											1				2		3			2					10
buah			1																		1		1				1	1					5
mengikut	1							1										1	1					1	1			1					7
datang							1	1										1									2						5
kurang											1							1	1								1	1					5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
pukul		1										1							1									1	1				5
sekiranya									1		1					1			1					1	3								8
kuasa									1										1				1		1		1	1					6
melakukan	1					1													2						1		2	1					8
barubaru			2								2																2	2					8
hidup			1					1													2			1					1				6
perlawana n		1				2					1	1											1		1		3						10
sektor							1		1													1				1			1				5
kejayaan						1								1			1		1						1		3						8
menunjuk kan	1					3								1					1			1			1		1	1	1				11
lelaki												2							1					1	1		1						6
kirakira											2								2					2	2								8
sri									1		1													1									3
laporan		1				1					1	1															2		1				7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
tanah						1	1														1						2						5
menyataka n	1					1	1									1			1						2		2						9
negaraneg ara						2					2									2					4		2						12
diberi			1					1			1													2	1								6
bandar			1			1		1																			2						5
manusia	1					1			1								1					1		1	1		1	1					9
abdullah			1					1				2									1						2	1					8
hasil									1			1									1			1			1						5
pasti		1					1		1															1			1						5
mac	1												1														1						3
menyebab kan	1		2			1												1	1						2		2						10
kecil												1	1						1					1	1								5
sebaliknya			1						1			1				1						1		1	2		1						9
meningkat kan	1					2		1										1	2					1	1		2						11

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
agama	1																			1					1		2						5
buat			1					1															1				1	1					5
mencari	1					1					1		1											1	1		1						7
awam	1																						1				2						4
menurutnya	1					1		1			1					1									2			1	1				9
memang	2																	1								1	1						5
pulau		1										1																1			1		4
tetap			1				2																		1		1						5
masingmasing	2								2									2						2			2						10
bermula	1		1									1													2			1					6
skuad								1	1										1				1				1	1					6
lepas		1							1			1													1		1						5
pengurusan		1				1			1		1							1							1		1	2					9
golongan						1						1						1		1							1		2				7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
penduduk		1				1		2														1			1			1	1				8
jawatan						1	1							1									1				3						7
mudah	1							1													1						1	1					5
ujarnya											1			1		1									1		1	1					6
selangor									1			1						1							1		1		1				6
mesyuarat	1							1			1				1								1		1		2	1					9
rendah						1		1			1										1				1		1						6
kerjasama	1								1		1			1					1						3		1						9
jepun		1				1								1											1			1					5
berdasarkan			1			1		1	1		1								1						1		3						10
nama	1					1																			1		1						4
maka	1																		1						1		1						4
kedudukan						1		2											2						1		1	1	1				9
terlibat			1				1	1			1													1	1		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
undangundang						2		2										2									2	2					10
hubungan			1			1												1			1						1	1	1				7
bernama	1		1			1					1														2		1						7
yakin						1											1		1					1			1						5
tentera						1	2				1														3								7
pengguna		1				1												1		1					2			1					7
tahu							1														1						1	1					4
kembali	1		1									1							1					1	1		1						7
anakanak						2																2					4						8
barangan			1			1					1							1									3						7
alam	1											1															1						3
agar																				1							2						3
digunakan						2		1										1		1					1		1	1					8
nilai						1						1												1						1			4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
terbabit			2				1	1			1													1	1		1						8
ingin						1												1						2									4
terlalu							1				1	2													1		1	1					7
utusan						1	1		1																		1	2					6
terbuka			1				1				1								1						2			1					7
mula	1											1													1			1					4
johor														1							1								2				4
mohamad	2							1													1						2	1					7
menambah	2		1			1															1				1		2						8
melibatkan	1		1			1		1				1							1					1	1		2						10
terutama	1						2				1														2		1	1					8
cukup			1										1						1									1	1				5
tahap			1				1														1						2						5
muda	1							1																	1			1					4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
muncul	1					1						1	1															1	1				6
pagi		1																		1				1			1						4
diberikan			1			1		1			1								1					2	1		1						9
tiada							1	1									1							1	1		1						6
sen						1			1																	1							3
pinjaman	1	1				2								1										1			2						8
acara											1		1												1		2						5
memerluk an	2					1					1	1							1						2		1	1					10
tak							1															1					1						3
perkemban gan	1	1	1			1					1							1	1						2		2						11
kesan						1			1										1						1		1						5
sabah			1						1												1						2						5
as																																	0
proses		1							2		1															1			1				6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
samping	1	1							1									1						1			1						6
jadi								1						1										1			1						4
tenaga						1	1													1					2		1						6
piala		1										1					1							1	1		1						6
bapa		1	1																						1		1						4
menyertai	1						1									1								1	2		1						7
guru											1									1								2					4
hampir	1	1																			1					1	1						5
enam	1					1																			1		1						4
kampung	1	1																1	1								1		1				6
petang		1					1											1							1		1						5
muzik	1									1							1					1		2				1					7
menyediakan	1					1		1								1	1		1					1	3		1						11
biasa			1						1								1							1	1		1						6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
sikap			1						1										1					1			1						5
seni						1			1															1	1								4
khas									1												1						1						3
kemudahan	1					1		1													1				1		2	1					8
bantuan			1			2	1																1				2	1					8
pelaburan		1	1			1					1	1													1		2		1				9
sokongan						1			1									1	1								1		2				7
perdagangan		1				1		1			1							1		1					1		3						10
pekerja		1									1			1					1						3								7
meminta	2					1	1																	1	2								7
mencapai	1	1				1							1												1		1			1			7
berhubung			2								1							1			1				1			1	1				8
perak		1									1											1				1	1						5
bola			1									1															1		1				4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
rasa									1		1														1		1						4
cuba			1										1												1			1					4
perhatian		1				1	1				1						1				1			1	1		2						10
segi									1											1				1	1								4
jawatanku asa						1	1		1					1					1				2		1		4	1					13
sdn						1		1	1		1						1							2	1		1						9
sebahagian			1			1			1								1			1	1			1	1		3						11
agak																				1		1					2						4
media	1							1									1							2			1						6
bekerja			1								1			1					1						3								7
tertentu						1	3																		2			1					7
badan			1			1		1																			2						5
aziz										2														1			1						4
kos									1										1										1				3

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
sedikit								1	1									1						1			1						5
sidang																																	0
sepanjang						1			1					1				1							1		2						7
pengerusi		1							1		1							1						1	2			1					8
sekitar							1		1										1					1	1		1						6
pejabat		1	1											1											1		2						6
musim	2								1																	1		1					5
tugas							1		1											1							1	1					5
asean						1			1								1							1			2						6
dianggap			1					1										1		1				1			2						7
persembahan	1	1	1			1			1												1				2		2						10
kaum	1																		1				1				1		1				5
langsung									1			1						1									1		1				5
sentiasa							1		2								1							1	2		1						8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
ibrahim	1		1								1										1			1		1	1						7
kereta							1				1								1						2	1							6
memboleh kan	2		1			1						1							1		1				1	1	1		1				11
peminat	1	1				1		1																1	1		1						7
dewan						1		1															1			1	1						5
segala									1			1								1					2		1						6
remaja	1										1			1											2		1						6
aktiviti					1		2															1		3			1						8
selatan						1	1		1			1													1		2						7
sesuai									2														1		1			1		1			6
rancangan						2					1		1					1									3						8
menjadika n	1					2		1						1					1					1	1		2						10
anda						1		1																			2						4
peguam	1	1																		1			1		1		1	1					7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
barat			1					1			1																2						5
berharap			2								1										1				1		2						7
bentuk			1			1	1															1			1				1				6
pengalaman	1	1				1						1						1							1		3						9
institusi						1	2		2								1							4				1					11
pengajian		1				1								1			1	1						1	1		2						9
hakim	1																		1		1			1			1						5
isteri							1		1		1													2	1								6
sumber	1		1						1																1			1					5
berkaitan			1			1	1				1						1		1					1	1		2						10
gagal												1								2							2						5
kemudian	1					1		1									1		1					1	1		1	1					9
kebanyakan			1			1										1			2						1		3						9
makanan	1					2													1								3						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
indeks						1		1	1													1		1		1							6
sains						1			2																					1			4
waktu							1															1	1				1	1					5
jelasnya									1			1		1		1									2		1						7
pelakon		1				1						1							1						1		1		1				7
bakal			1									1							1								2						5
artis							1		1		1													1			1						5
menjalankan	1					3						1		1					1						1		3						11
australia							1		1		1	1					1							1			2		1				9
dasar								1	1		1																2						5
dikatakan						1	1	1											2					1			3						9
swasta							1		2														1				2	1					7
krisis									2		1								1					2									6
berikutan			1			1	1				1								1					1	1		1	1					9

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
hal												1									1						1						3
sebab			2						1																1		1						5
setiausaha							1		2								1				1			1	1		3	1					11
nasional						2						1			1									1			2		1				8
perubahan		1	1			1					1										1				1		2	1					9
menjelaskan	1					2			1			1		1					1						2		2						11
pemberita	1	1	1				1				1														3								8
hak																					1	1					1						3
pinang		1				1												1						1			1						5
keselamatan	1					1	1		1			1							1						2		3						11
komputer	1	1					1												1						1			1	1				7
sepak		1							1													1				1	1						5
cadangan						1		1					1					1									3						7
penuh		1				1															1				1				1				5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
buku			1																1									2					4
kepentingan		1				2	1											1	1					1	2		1						10
sukar									1		1								1								1	1					5
seseorang									2		1							1							2		1		1				8
arah											1										1						2						4
diambil	1		1					1				1												1		1	1						7
gus									1											1								1					3
ketiga							1												1	1				1	2								6
peranan		1				2					1																2						6
pertemuan	1	1				1	1				1												1		2		1	1					10
suami	1								1														1	1			1	1					6
jurulatih							1				1	1		1							1					1	1	2					9
pertandingan		1				2	1	1			1							1						1	1		2						11
raja											1			1											1		1						4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
emas	1								1																1		1						4
wan						1																	1				1						3
pelabur		1	1									1													1		1		1				6
cabaran			1			1					1		1														3						7
produk		1						1			1											1					1		1				6
kehidupan		1				1		1											1		1			1	1		1	1					9
meneruskan	1					2			1		1								1						2		1	1					10
setakat							1	1	1										1						1		2						7
jan						1								1											1								3
operasi		1							1		1													1	1		1		1				7
soal									1			1															1		1				4
jauh														1							1						1		1				4
sering									1		1							1							2								5
perjanjian		1				2					1			2			1							1	1		2						11

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
dana						1		1																			2						4
akibat			1					1											1					1			2						6
sembilan	1		1			1			1			1												1	1		1						8
keperluan		1				1					1	1							1				1		2		1	1					10
kota							1												1						1				1				4
pembinaan	1	1	1			2																1		1	1		2						10
bertindak			1			1	1	1			1											1		1	1		1						9
pelajaran		1				1					1	1		1											1		3						9
terengganu						1	1				1							1		1					3			1					9
dilaksanakan						2		1	1			1							2					1	1		3						12
nanti						2	1																	1			1						5
kelantan						2	1					1							1						1		2						8
pandangan		1				1		1										1									3						7
apakah		1																	1		1				1		2						6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
minit	1					1	1																	1	1								5
jam	1													1													1						3
april		1									1	1													1		1						5
memandan gkan	2					2		1										1	1						1		3						11
juara											1			1									1		1		1	1					6
berusaha			1						1		1										1				3			1					8
pelanggan		1				1						1						1		1					1		2						8
kamu	1																		1								1	1					4
sejarah									1		1			1							1				1		2						7
dahulu								1				1									1						1	2					6
ditemui	1						1	1															1	2	1			1					8
membeli	2	1										1												1	2								7
sambil	1		1						1			1												1	1								6
surat							1		1		1																1	1					5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
kajian						1								1					1					1			2						6
pihaknya		1														1			1		1				2		1						7
dagangan						1		1										1		1							3						7
ringgit							1				1							1		1				1	1								6
semangat	1							1	1									1							1		2						7
kakitangan						1	1											1	2					1			3						9
memilih	2											1									1			1	2								7
kegiatan						1	1										1		1	1				1	1		2						9
menghasil kan	1					1			1			1						1	1		1			1	1		2						11
kejadian						1		1						1			1		1					1	1		2						9
januari						1					1			1									1	1	1		1	1					8
keluar												1							1				1		1		1	1					6
diterima	1						1	1			1													2	2								8
mengalami	2											1						1						1	1		2						8
sambutan	1		1			1	1		1																		2	1					8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
menguran gkan	1					1					1							2	1						1		2	1					10
jenis						1			1					1										1	1								5
suatu							1		1														1				1	2					6
budaya			1					1									1								1		1	1					6
khususnya									2							1			1		1				1			2					8
feb			1	1							2												1	1	1		1	1					9
kanakkana k						2													2			2					4						10
tanggungja wab			1				1							1				2		1			1				3		1				11
ismail	1								1			1												1		1	1						6
kesihatan						1	1		1										1		1				1	1	2						9
masuk	1								1													1					1		1				5
bagaimana	1		1			1														1				1	1		3						9
rantau						1	1				1																2	1					6
percaya		1									1		1				1								2		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
thailand						1	1					1													1		1						5
sepatutnya		1					2		1							1									2		1	1					9
tambahnya	1		1				1									1					1				1		2						8
konsep		1				1			1										1						1				1				6
unit						1	1										1							1				1					5
mengeluar kan	1					1					1	1						1	1				1		2		2	1					12
kemungki nan	1					2												1	2					1	1		1	1					10
urusan						1			1		1																1	2					6
manamana	2					2																			2		2						8
sarawak									1		1											1	1		1		2						7
tangan						1	1											1									2						5
bukit			1				1												1							1		1					5
kegawatan						1	1												1	1			1		1		3						9
terbaru			1				1				2														1		1	1					7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
pengurus		1							1		1							1							1			2					7
menyaksikan	1					1			1							1			1			1		1	1		2						10
melaksanakan	1					2			1			1							2						2		3						12
modal	1							1				1															1		1				5
dilihat							1	1				1									1			2			1						7
lagulagu												2								2							2	2					8
tujuan						1	1							1									1				1	2					7
hendak						1		1													1	1			1		1						6
bahan			1			1															1						2						5
sebenar			1			1			1																2		1						6
pertumbuhan	1	1	1			1	1				1										1				1		1	1	1				11
popular		2										1													1			1	1				6
terutamannya	1						2				1					1									3		1	1					10
naik						1											1					1				1	1						5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
dulu								1				1																2					4
pertubuhan		1	1			1	1				1										1				1		1	1	1				10
tengah							1											1			1				1		1						5
keuntungan						2	1											1	1						1		1	1	1				9
china						1							1											1	1								4
penggunaan		1				2												1	1						1		2	1					9
ditanya							1	1								1								1	1		1						6
pahang		1																1			1						2						5
kenaikan						1													2						1	1	2						7
calon						1						1	1														1		1				5
hotel							1					1									1				1				1				5
peningkatan		1				2	1											1	1						1	1	2						10
sosial									1			1			1										1				1				5
kanan						2													1								2						5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
ali												1												1			1						3
tambah	1		1				1														1						2						6
mengulas	1								1			1						1							1		1	1					7
stadium	1						1	1	1								1							1	1		1						8
mesti	1						1		1															1		1							5
sedia								1	1								1							1	2								6
ogos									1											1					1				1				4
bn			1			3			1		1	1			1									1	2		3						14
cina						1							1											1	1								4
zaman	1					1				1																	2						5
menjelang	1					1						1		1				1							2		1						8
bersedia			1					1	1								1							1	3								8
permohonan	1	1				2															1				1		1		2				9
panjang		1				1								1				1									2						6
sungai									1									1						1			1	1					5

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
kursus									2										1									2					5
dijalankan						2		1				1		1					1					1			3						10
kemenangan	1					2												1	1						2		2						9
mulai	1											1																1		1			4
melaka	1											1							1						2		1						6
jatuh							1							1							1						1		1				5
anugerah						1					1									1	1				1		2	1					8
balik			1									1										1				1	1						5
asas									2																		2						4
pas		1							1																		1						3
timur	1						1																	1					1				4
pulang		1										1						1									1	1					5
sepenuhnya		1				1			1							1					1				3				1				9
shah															1						1						1						3

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
persekutuan		1				1	1		1										1				1		2		1	2					11
julai												1		1														1		1			4
jualan						1						1		1									1				2	1					7
sangat								1	1									1									2						5
bermakna	1		1			1													1						2		1						7
bangsa			1						1									1							1		1						5
disember	1		1					1		1														1	2								7
pemilihan	1	1				1						1									1			1	1	1	1						9
berkuasa			1						1										1				1		2		1	1					8
melahirkan	1					1						1							1		1				1	1	2						9
setahun						1	1		1												1				1		1		1				7
meletakkan	1					1	1					1							1			1			2		2						10
akhirnya																1					1				1	1	1						5
inilah						1						1									1			2			1						6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
disebabkan			2			1		1	1										1					1	1		2						10
itulah							1					1									1			1			1	1					6
ikatsyarikat								2			1				1				2					2			3						11
esok									1													1				1			1				4
memiliki	2											1							1					3	1								8
komanwel	1					1						1							1				1		2				1				8
minat	1					1		1																1			1						5
papan		2				1																					2						5
internet						2	2																	1	2								7
umum	2																											1	1				4
eropah		1									1										1					1	1		1				6
bila			1									1												1	1								4
barisan			1			1			1		1															1	2						7
demi	1							1																1	1								4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐ u	oɪ	Sum
mengatasi	1						1		1									1						1	1		2						8
ilmu	1											1													1			1					4
rakan						1					1								1								2						5
september	1	1	1				1		1																3								8
kedah								1											1		1				1		1						5
perbincangan		1	1			2					1		1					1						1	1		2						11
eksekutif				1			1			1									1	1					3								8
peserta		1					1		1																3								6
keseluruhan						1			1		1	1							1		1				2		1	1	1				11
sampai	1	1							1															1			2						6
jangka														1				1	1						1		1						5
menawarkan	1					2													1				1		1		3						9
kira											1								1					1	1								4
segera									1		1									1					3								6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ε	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
aspek		1							1													1			2								5
terakhir							1				1										1				1	1	1						6
pusingan		1				1			1									1						1			1	1					7
bebas			2						1																	1	1						5
membuka	2		1																1						2			1					7
arahan						1					1										1						3						6
sumbangan	1		1			1			1									1									2	1					8
kempen	1	1				1													1						2								6
serangan						1			1		1							1							1		2						7
hospital		1					1		1			1									1			1			1		1				8
hadapan		1				1		1													1						3						7
pingat		1						1										1						1			1						5
gol												1								1									1				3
pengeluaran		1				1					1	1						1					1		2		2	1					11

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
penyakit		1					1									1			1						1	1	1						7
seterusnya							1		2		1					1									3				1				9
serius									2		1						1							1	2								7
jakarta							1				1			1					1								3						7
berpendapat		1	1			1		2																	2		2						9
naib			1			1											1								1		1						5
mengakui	1																	1	1				1	1			1	1					7
agensi						1			1											1				1	1	1							6
berusia			1						1		1						1							1	2			1					8
faedah				1				1													1					1	2						6
jangan						1								1				1									2						5
tambahan	1		1			1	1														1						3						8
memulakan	2					1						1							1						1		2	1					9
drama	1							1			1																2						5
keyakinan						2											1		2						1	1	2						9

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐ ɪ	ɐ u	o ɪ	Sum
liga												1								1				1			1						4
bergantung			1			1	1											1		1					1		1		1				8
prestasi		1					1		2		1													1	1		1						8
tumpuan	1	1				1	1																1				1	2					8
mewujudkan	1					1		1						1					1				1		1		1	2					10
tegas							1		1											1					1		1						5
menganggap	1	1																2		1					1		2						8
peraturan		1				1	1				2														1		2		1				9
berbeza			2							1															2	1							6
rahman	1					1					1										1						2						6
cuma	1												1												1			1					4
tujuh							1							1							1							1	1				5
terdiri							1	1			2													2	1								7
bersetuju			1				1		1					1											2			2					8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
sultan						1	1		1			1															1	1					6
peniaga		1				1											1			1				1	2		1						8
suasana						1			2														1		1		2	1					8
abu			1																								1	1					3
menggalak kan	1					1												1	1	1		1			1		2						9
menentang	1					2	1											1							2		1						8
berat			1				1				1														1		1						5
perasaan		1				1			1		1											1			1		3						9
tangga							1											1		1					1		1						5
kontrak						1	1				1								1			1					1		1				7
menolak	1					1						1										1			1		1		1				7
membukti kan	2		1			1	1												2					1	1		1	1					11
minyak	1															1						1		1			1						5
membina	2		1			1																		1	2								7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
dikeluarkan						1		1			1	1							2				1	1	1		2	1					12
siasatan						1	1		2								1							1			3						9
menjalani	1					2						1		1										1	1		2						9
bertujuan			1			1	1				1			1									1		1		1	2					10
berita			1				1				1													1	2								6
memenuhi	2					1															1			1	2			1					8
dikenali						1		1				1							1					2	1		1						8
berminat	1		1			1	1																	1	1		1						7
rakaman	1					1					1								1								3						7
menangani	1					2												1						1	1		2						8
jaya														1			1								1		1						4
mei	1																							1	1								3
mencatat	1					2	2						1						1						1		3						11
soalan						1			1			1															2		1				6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
pentadbiran		1	1			2	1	1			1													1	1		2						11
faktor				1			1												1								1		1				5
hadir								1													1				1		1						4
pendapatan		2				2	1	1																	1		3						10
saja									1					1											1		1						4
allah												2									1						1		1				5
sehubungan			1			1			1									1			1				1		1	2					9
rasmi	1								1		1													1			1						5
harapan		1				1					1										1						3						7
berjumlah	1		1								1	1		1							1				1		1	1					9
kesemua	1								1										1				1		3			1					8
dijadikan						1		2						1					1					2			2						9
individu					1	1		2																3				1					8
akta							1												1								2						4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
kuat								1											1				1				1	1					5
wilayah												1					1				1		1	1			2						7
tentu						1	2																		1			1					5
dikenakan						2		1											2					1	2		1						9
maju	1													1													1	1					4
sebelumnya	1		1						1			1				1									3				1				9
bangunan			1			2												1									2	1					7
tun						1	1																					1					3
ekoran						1					1								1							1	1		1				6
lapan		1				1						1															2						5
menegaskan	1					2			1										1	1					2		2						10
sesetengah							1		2									1			1				3		1						9
penyelidikan		1				1		1				1				1			1					1	2	1	1						11
pesawat		1					1		1														1		1		2						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
india						1		1									1							2			1						6
kenderaan						2		1			1								1			1			2		2						10
timbul	1		1				1					1												1					1				6
mengekalkan	1					1						1						1	2						2		2						10
berlangsung			1			1			1		1	1						1							1		1		1				9
contoh						1	1						1								1								2				6
permintaan	1	1				2	1				1											1		1	1		2						11
dirinya								1			1					1								2	1								6
permainan	1	1				2					1						1								2		2						10
sebagainya			1						1							1				1				1	2		2						9
ujian						1								1			1							1			1	1					6
menimbulkan	2		1			2						1							1					1	1		1	1					11
umat	1							1																			1	1					4
peruntukan		1				2	1				1								1			1			1		1	1	1				11

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
inggeris									1		1							1		1				2	1								7
atlet							2					1														1	1						5
mewakili	1											1							1				1	2	1		1						8
hassan						1			1												1						2						5
kelab			1									1							1						1		1						5
berubah			2								1										1				1		1	1					7
multimedia	2						1	1				1					1							2		1	1	1					11
taman	1					1	1																				2						5
tinggal							1					1						1		1				1			1						6
menentukan	1					3	1												1						2		1	1					10
sempena	1	1				1			1																3								7
adakah								1											1		1				1		2						6
mangsa	1								1									1							1		1						5
sesi									2															1		1							4

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
tuan						1	1																1				1	1					5
cawangan						1							1					1					1				3						7
daerah								1			1										1					1	2						6
laut							1					1											1				1		1				5
separuh		1							1		1										1				1		1		1				7
berakhir			1								1										1				1	1	1						6
memenangi	2					1												1						1	2		1						8
bukanlah			1			1						1							1		1						2	1					8
ujar											1			1													1	1					4
import	1	1					1																	1					1				5
seharusnya									2		1					1					1				2		1	1					9
mendapati	1	1				1	1	1																1	1		2						9
parlimen	1	1				1					1	1												1	1		1						8
beraksi			1						1		1								1					1	1		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
kemajuan	1					1								1					1				1		1		2	1					9
persidangan		1				1		1	1		1							1						1	1		2						10
berkesan			1			1			1										1						1		1						6
belajar			1									1		1											1		2						6
lawatan						1	1					1											1				3						7
positif		1		1			1			1														2					1				7
korea											1						1		1					1			1		1				6
siti							1		1															2									4
lebuh			1									1									1				1				1				5
menghantar	1					1	1											1			1				1		2						8
sekadar								1	1										1						1		2						6
tawaran						1	1				1												1				3						7
perubatan		1	1			1	1				1														1		2	1					9
sedangkan						1		1	1									1	1						1		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
tiba			1				1																	1	1								4
hong																		1			1								1				3
kategori							1				1								1	1				1	1		1		1				8
buruk			1								1											1						1	1				5
siapa		1							1								1							1	1		1						6
mohamed	2							1													1						2		1				7
kawalan						1						1							1				1				3						7
belakang			1									1						1	1						1		2						7
diperlukan		1				1		1			1	1							1					1	1		1	1					10
matlamat	2						2					1															3						8
pelumba	1	1	1									1													2			1					7
wajar											1			1									1				2						5
bayaran			1			1					1						1										3						7
membayar	2		1														1								1		2						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
mengikuti	1						1											1	1					1	1			1					7
bursa			1						1		1																1	1					5
menyokong	1															1		1	1						1				2				7
kekuatan							1												2				1		1		2	1					8
berterusan			1				2		1		1														2		1		1				9
pakar		1																	1								2						4
ditutup			1				2	1																1				1	1				7
menunggu	1					1												1		1					1			2					7
udara								1			1														1		1	1					5
sasaran						1			2		1																3						7
alasan						1			1			1															3						6
menjejaskan	1					2			1					2					1						2		2						11
khidmat	1							2													1			1			1						6
bertambah	1		2				1				1										1				1		2						9

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐ ɪ	ɐ u	o ɪ	Sum
terkenal						1	1				1	1							1						2		1						8
ruang											1							1					1				1	1					5
keupayaan		1				1											1		1			1			1		3	1					10
kasih									1										1		1				1		1						5
penonton		1				3	1																		1				2				8
pencapaia n		2				2							1				1								1		3						10
bekalan			1			1						1							1						1		2						7
hadiah								1									1				2			1			2						7
sesebuah			1						2												1		1		2		1	1					9
kaedah								1											1		1					1	2						6
makan	1					1													1								2						5
jenayah						1								1			1				1				1		2						7
turun						1	1				1																	1	1				5
cemerlang	1											1	1					1							2		1						7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
dakwaan						1		1														1	1				3						7
kertas							1		1		1								1						1		1						6
bumiputer a	1	1	1				1				1													1	2			2					10
mengumu mkan	3					1												1	1						1		1	1	1				10
keterangan						1	1				1							1	1						2		2						9
agung																		1		1							1		1				4
hukuman	1					1													1		1						1	2					7
rasuah									1		1										1		1				2	1					7
pendekata n		1				2	1	1											1						2		2						10
korporat		1					1				1								1						1		1		1				7
laluan						1						2											1				2	1					7
pertanian		1				2	1										1							1	1		2						9
wujud								1						1									1					2					5
karya											1						1		1								2						5
pergerakan		1				1					1								1	1					2		2						9

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
membabitkan	2		2			1	1												1						2		2						11
memasuki	2								1										1					1	1		1	1					8
batu			1				1																				1	1					4
kukuh																			2		1							1	1				5
benar			1			1																			1		1						4
penulis		1				1			1			1													1	1		1					7
perbicaraan		1	1			1					1		1									1		1	1		3						11
bercakap		1	1										1						1						1		2						7
menyelesaikan	1					1			1			1				1			1					1	3		2						12
mati	1						1																	1			1						4
mengetahui	1						1											1			1		1	1	2		1	1					10
bilik			1									1										1		1		1							5
paras		1							1		1																2						5
pertahanan		1				2	1				1										1				1		3						10

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
lim	1											1												1									3
suka									1										1						1			1					4
wakil												1							1				1			1	1						5
terletak							2				1	1										1			2		1						8
belia			1									1					1							1	2								6
tuduhan						1	1	1													1						1	1	1				7
nak						1																1					1						3
berbuat			2					1															1		1		1	1					7
bersatu			1				1		1																1		1	1					6
dipilih		1						1				1									1			2		1							7
peribadi		1	1					1			1													2	1		1						8
kemahiran	1					1					1								1		1				1	1	2						9
bermain	1		1			1					1														1	1	1						7
lainlain						2						2													2		2						8

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
strategi							2		1		1			1										1	1		1						8
kehadiran						1		1			1								1		1				1	1	2						9
disediakan						1		2	1								1		1					2	1		2						11
kelmarin	1					1					1	1							1						1	1	1						8
terbesar			1				1		1																2		1						6
milik	1											1										1		1		1							5
augustine						1	1		1											1				1			1	2					8
utara							1				1														1		1	1					5
sanggup		1							1									1		1							1		1				6
eksport		1					1		1													1			1				1				6
pengetahuan		1				1	1											1			1		1		2		2	1					11
diminta	1					1	1	1																2	1								7
sematamata	2						2		1																1	2	2						10
kabinet			1			1	1												1					1	2								7

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	**ə**	**ɛ**	**ɐ**	**u**	**o**	ɐɪ	ɐu	oɪ	Sum
perbelanjaan		1	1			2						1		1								1			2		3						12
profesional		1		1		1					1	1			1										3				1				10
misalnya	1								1			1				1								1	1		1						7
jun						1								1														1					3
perjalanan		1				2						1		1											1		3						9
razak										1	1											1					2						5
amanah	1					1															1						3						6
sedar								1	1																1		1						4
mendakwa	1					1		1														1	1		2		1						8
pekerjaan		1				1					1			1					1			1			1		2						9
kelas									1			1							1						1		1						5
penyertaan		1				1	1									1						1			2		2						9
mengatakan	1					1	1											1	1						2		2						9
bertemu	1		1				1																		2			1					6

Phoneme	m	p	b	f	v	n	t	d	s	z	r	l	ʧ	ʤ	ʃ	ɲ	J	ŋ	k	g	h	ʔ	w	ɪ	ə	ɛ	ɐ	u	o	ɐɪ	ɐu	oɪ	Sum
saksi									2													1		1			1						5
Oktober			1				1												1						1				2				6
Kesalahan						1			1			1							1		1				1		3						9
Rangkaian						1					1						1	1	1								3						8
Usahawan						1			1												1		1				3	1					8
Tabung			1				1											1									1		1				5
Filipina		1		1		1						1												3			1						8
count	343	215	227	8	3	497	286	206	299	9	270	233	26	88	9	40	64	152	282	78	172	75	85	371	799	72	1158	303	133	13	4	0	6520
%	5.3	3.3	3.5	0.1	0.0	7.6	4.4	3.2	4.6	0.1	4.1	3.6	0.4	1.3	0.1	0.6	1.0	2.3	4.3	1.2	2.6	1.2	1.3	5.7	12.3	1.1	17.8	4.6	2.0	0.2	0.1	0.0	100
%	12.0			0.2		27.6							1.9			1.6		7.9			3.8		1.3	17.9		18.9		6.7		0.3			100

APPENDIX C: Test lists for the MDTT

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
063	023	031	028	047	035	051	040
081	037	042	052	060	057	075	054
085	032	043	084	070	068	082	080
106	131	161	105	101	125	126	136
130	162	178	128	165	140	147	148
156	185	184	157	172	168	182	183
238	240	251	242	212	204	213	218
243	247	274	272	231	217	263	237
278	284	286	270	276	275	271	267
308	348	304	360	320	302	303	317
310	363	313	364	357	324	321	352
367	387	373	385	386	353	380	378
421	418	402	401	423	468	407	405
427	424	425	415	454	481	432	420
484	465	452	457	486	482	470	475
532	506	517	501	525	530	504	503
540	513	545	521	548	571	536	562
572	520	567	568	564	584	565	574
605	656	610	630	618	601	614	615
612	658	626	634	632	627	648	631
657	676	637	636	684	673	654	681
714	705	756	713	703	716	715	702
723	711	768	743	735	736	726	746
765	754	785	786	741	750	760	753
841	802	808	816	807	812	857	824
854	870	820	847	813	843	838	826
876	871	830	873	858	846	848	861

APPENDIX D: Test lists for MMST-AV in TSN

List 1

dia ambil semua kotak kecil
abang ada tujuh buku hitam
kami suka tiga pisau merah
abang dapat dua baju besar
ayah mahu enam kunci putih
ayah nampak empat mangkuk putih
kakak minta empat topi mahal
dia bagi satu bola lama
dia ambil tiga kotak lama
dia ada banyak buku lama
kakak ambil tujuh mangkuk mahal
kami nampak semua mangkuk kecil
ayah minta tujuh topi putih
nenek minta tujuh topi cantik
kakak suka empat pisau mahal
kita minta banyak topi lama
kami suka lima pisau putih
adik dapat tujuh baju mahal
dia ambil satu kotak baru
adik ada lapan buku hijau
saya ambil lima kotak baru
kakak suka lima pisau mahal
abang mahu banyak kunci lama
ibu nampak semua mangkuk merah
dia mahu lapan kunci cantik
kakak dapat lima baju mahal
adik mahu tujuh kunci mahal
kakak beri lapan meja mahal
kakak ada tiga buku mahal
ibu nampak lima mangkuk merah

List 2

nenek perlu enam lampu cantik
adik minta enam mangkuk hijau
kami mahu empat kunci hitam
saya ambil semua kotak baru
dia dapat empat baju hitam
adik nampak tujuh mangkuk hijau
nenek perlu lapan baju cantik
kita mahu dua kotak besar
dia bagi semua bola lama
kakak mahu tiga kunci merah
kita ambil enam kotak besar

List 9

abang dapat empat baju hitam
ayah nampak semua mangkuk kecil
saya nampak tujuh mangkuk mahal
dia ambil dua kotak besar
ibu ada banyak buku lama
abang suka lima pisau hitam
adik dapat empat baju hitam
abang nampak enam mangkuk hijau
ibu nampak satu mangkuk merah
ayah nampak banyak mangkuk lama
kakak mahu enam kunci hijau
kita dapat banyak baju lama
dia ambil tiga kotak merah
kami beri semua lampu kecil
saya ambil satu kunci baru
kami ada empat buku hitam
abang ada lima buku hitam
ibu ada lima buku putih
adik beri enam baju hijau
ayah beri dua meja besar
kita ada banyak buku besar
kakak suka tujuh pisau mahal
kami ambil lima kotak putih
nenek ambil lima kotak cantik
abang perlu enam lampu hitam
kita suka lapan pisau besar
adik suka enam meja hijau
ibu perlu tiga kunci merah
ibu ambil banyak kotak lama
saya suka satu topi baru

List 10

nenek perlu lima lampu cantik
kami ambil lapan kotak cantik
kami beri empat meja hitam
nenek dapat lapan lampu cantik
ayah beri dua meja putih
kita beri banyak meja lama
ayah dapat satu baju baru
ayah nampak tiga mangkuk putih
abang suka tujuh pisau mahal
dia ambil banyak pisau lama
nenek minta empat topi cantik

abang dapat tiga baju hitam
abang suka semua pisau hitam
kakak suka dua pisau besar
ayah nampak dua mangkuk putih
abang nampak empat topi hitam
kami ada lima buku putih
ibu nampak tiga pisau merah
adik nampak semua mangkuk kecil
dia minta satu topi baru
ibu dapat lima baju merah
adik dapat empat baju hijau
adik ambil enam kotak hijau
nenek suka lapan topi cantik
adik minta satu topi baru
kita mahu satu kunci besar
nenek ambil tiga kotak merah
abang ambil banyak kotak hitam
kakak suka tiga pisau merah
kakak dapat empat baju mahal

List 3

abang suka empat lampu hitam
kakak beri enam meja hijau
kami ambil empat kotak hitam
dia ada tiga buku merah
nenek ambil semua kotak kecil
saya dapat dua baju baru
ayah ambil tujuh kotak putih
kita dapat empat baju hitam
nenek beri lapan kunci cantik
ayah beri empat meja putih
adik nampak semua mangkuk hijau
kakak nampak lapan mangkuk mahal
dia suka empat pisau lama
nenek suka dua pisau besar
kami ambil dua kotak besar
nenek minta lima topi putih
kita nampak lapan mangkuk cantik
dia bagi banyak bola lama
kakak ambil tujuh meja mahal
kita suka tiga pisau merah
kakak nampak banyak mangkuk mahal
saya ambil dua kotak baru
abang nampak banyak mangkuk lama
ayah nampak empat mangkuk hitam
abang ada banyak buku lama
ayah suka dua pisau putih

kakak ambil lapan kotak mahal
kakak minta tujuh mangkuk mahal
saya minta tiga topi baru
saya nampak lapan mangkuk cantik
saya ada empat buku hitam
kita suka banyak pisau besar
ibu minta tiga pisau merah
kita beri satu meja besar
adik mahu enam kotak hijau
ayah suka banyak pisau putih
dia ambil tujuh kotak lama
kita bagi semua bola kecil
kita nampak enam mangkuk besar
kami nampak tiga mangkuk kecil
kami beri enam meja hijau
adik beri semua meja kecil
kakak suka banyak pisau lama
nenek ambil dua kotak cantik
ayah ambil tiga kotak merah

List 11

abang ambil empat kotak hitam
abang ambil tiga kotak hitam
kakak minta lapan topi cantik
adik beri tujuh meja hijau
nenek nampak empat mangkuk hitam
ayah nampak enam mangkuk hijau
kita beri semua meja kecil
adik beri lapan meja hijau
kami nampak semua pisau kecil
kakak minta lima topi putih
ibu dapat banyak baju lama
adik ada semua buku kecil
kakak ada enam buku mahal
adik minta tiga topi hijau
nenek nampak lima mangkuk cantik
abang nampak satu mangkuk hitam
ayah mahu empat kunci hitam
ibu suka satu pisau baru
dia suka lapan pisau lama
abang minta empat topi hitam
nenek nampak semua mangkuk kecil
kakak minta tujuh topi mahal
saya ambil enam kotak baru
kami nampak enam mangkuk kecil
kakak suka tujuh meja mahal
kita suka dua pisau besar

kami ada enam buku kecil
abang mahu tujuh kunci hitam
nenek perlu semua lampu kecil
saya suka tujuh pisau baru

List 4

ayah dapat dua baju putih
saya minta satu topi baru
dia nampak empat mangkuk hitam
ayah dapat enam baju hijau
adik beri banyak meja lama
saya suka dua pisau baru
dia minta tiga topi merah
ibu ada empat buku hitam
dia ada semua buku lama
kami ada banyak buku kecil
saya nampak lapan mangkuk baru
ayah ambil empat kotak hitam
adik ambil tujuh kotak hijau
ibu perlu tujuh lampu merah
adik ada lima buku hijau
kita ambil lima kotak besar
adik beri lima meja putih
nenek mahu tujuh kunci mahal
ibu dapat satu baju merah
saya suka tiga pisau merah
ibu ambil empat kotak hitam
abang mahu empat mangkuk hitam
dia nampak enam mangkuk lama
nenek suka tiga pisau cantik
ibu minta satu topi merah
adik nampak lapan mangkuk cantik
adik nampak satu mangkuk baru
saya suka satu mangkuk baru
kami beri semua meja kecil
nenek minta enam topi cantik

List 5

ayah dapat satu baju putih
kami ada enam buku hijau
ayah beri lapan meja putih
kami perlu semua lampu kecil
kami nampak dua mangkuk besar
adik ambil tiga kotak merah
adik ada lima buku putih
ayah ada banyak buku lama
kita dapat lima baju besar
dia ambil dua kotak lama

adik minta tiga topi merah
saya nampak satu mangkuk baru
dia beri enam meja hijau
kita suka dua kunci besar

List 12

abang ada tiga buku hitam
nenek dapat enam baju cantik
ibu ada enam buku merah
abang suka empat pisau hitam
ayah minta lapan topi cantik
dia minta banyak mangkuk lama
kakak perlu semua lampu mahal
ayah ambil enam kotak hijau
kakak mahu tujuh mangkuk mahal
saya minta lima topi putih
kita beri banyak meja besar
adik nampak lima mangkuk putih
kita ambil dua pisau besar
ibu suka tiga pisau merah
saya nampak semua mangkuk kecil
ayah ambil semua kotak kecil
kakak nampak lima mangkuk putih
kita ambil banyak kotak lama
saya ambil satu kotak baru
saya ada satu buku baru
kakak ada tiga buku merah
kami beri dua meja besar
abang ambil lima kotak hitam
dia dapat banyak baju lama
ayah minta tiga topi merah
kakak ada empat buku hitam
dia suka banyak meja lama
abang minta empat mangkuk hitam
ibu minta tiga topi merah
ayah perlu semua lampu putih

List 13

abang ada dua buku besar
kakak minta tiga topi mahal
abang ambil lapan kotak cantik
adik nampak satu mangkuk hijau
kakak mahu tujuh kunci mahal
kakak perlu empat lampu hitam
dia suka banyak topi lama
abang ada dua buku hitam
dia ada lima buku putih
nenek mahu lapan kunci cantik

kami bagi dua bola besar
abang nampak dua mangkuk besar
ayah suka tiga pisau merah
kita mahu satu kunci baru
abang dapat enam baju hitam
dia beri semua meja kecil
adik bagi lapan bola cantik
abang perlu lima lampu putih
kakak ambil satu kotak baru
ayah minta satu topi baru
nenek suka lapan kunci cantik
dia suka semua pisau lama
kita dapat tiga baju merah
adik ambil enam topi hijau
kita beri enam meja besar
dia bagi dua bola lama
nenek perlu lapan lampu cantik
saya dapat tiga baju baru
kita nampak semua mangkuk kecil
nenek beri enam meja cantik

List 6

nenek beri enam meja hijau
nenek mahu dua kunci besar
ayah ada banyak buku putih
dia nampak dua mangkuk lama
nenek dapat enam baju hijau
abang nampak enam mangkuk hitam
ibu suka lima pisau merah
kakak suka satu pisau mahal
abang minta satu topi baru
ibu nampak banyak mangkuk merah
ayah nampak lima pisau putih
dia beri empat meja lama
adik beri satu meja hijau
kakak mahu tujuh topi mahal
nenek suka semua pisau cantik
saya minta dua topi baru
kita nampak banyak mangkuk besar
nenek nampak tiga mangkuk merah
adik nampak banyak mangkuk hijau
ayah ambil banyak kotak putih
ayah ada tiga buku merah
adik dapat enam baju hijau
nenek nampak banyak mangkuk cantik
saya ambil satu pisau baru
dia beri lima meja putih

kami nampak lapan mangkuk kecil
nenek nampak dua mangkuk cantik
nenek minta satu topi cantik
dia suka tujuh pisau lama
dia nampak satu mangkuk baru
saya suka dua pisau besar
kakak ambil banyak kotak mahal
abang ambil empat kunci hitam
saya ambil lima kotak putih
kita ambil dua kotak besar
adik minta enam kotak hijau
ayah ada enam buku hijau
abang perlu empat baju hitam
ayah dapat tujuh baju mahal
kakak ambil enam kotak mahal
saya perlu enam lampu hijau
saya ambil banyak kotak lama
kakak beri enam meja mahal
kita minta tiga topi merah
ibu perlu lima lampu putih

List 14

nenek dapat empat baju hitam
saya suka satu kotak baru
kami perlu lima lampu putih
nenek nampak tujuh mangkuk mahal
ayah ambil lima kunci putih
dia nampak satu mangkuk lama
saya nampak tujuh mangkuk baru
ibu suka banyak pisau merah
ibu nampak lapan mangkuk cantik
kakak beri lima meja putih
ayah beri tujuh meja putih
kami ada dua buku kecil
adik beri enam lampu hijau
kami nampak lima mangkuk kecil
kita dapat satu baju baru
kakak nampak empat mangkuk hitam
abang nampak semua mangkuk hitam
adik dapat lima baju hijau
kami ambil lima kotak kecil
kita suka lima pisau putih
ayah mahu lapan kunci cantik
adik mahu lapan kunci cantik
kami ambil dua kotak kecil
abang dapat lapan baju cantik
abang nampak empat kotak hitam

nenek suka lima pisau putih
kami perlu enam lampu hijau
adik ambil lapan kotak cantik
ibu nampak dua mangkuk merah
abang ada banyak buku hitam

List 7

kakak bagi lapan bola cantik
ayah nampak dua mangkuk besar
dia suka dua pisau lama
ibu beri satu meja baru
ayah mahu empat kunci putih
ayah ada tujuh buku putih
ibu ada semua buku merah
abang ada empat buku hitam
kakak nampak satu mangkuk baru
kami nampak satu mangkuk baru
kakak dapat empat baju hitam
kami nampak enam mangkuk hijau
ayah dapat enam baju putih
kita ambil dua mangkuk besar
nenek suka lapan pisau cantik
kakak mahu tujuh meja mahal
nenek ambil satu kotak baru
ibu ambil lima kotak putih
dia dapat enam baju hijau
kakak nampak satu mangkuk mahal
ibu ada semua buku kecil
kita suka semua pisau kecil
ayah minta tiga topi putih
kakak perlu lapan lampu mahal
nenek ada lapan buku cantik
dia suka tiga pisau lama
kami ambil tiga kotak kecil
kita nampak satu mangkuk baru
nenek suka lapan lampu cantik
adik beri satu meja baru

List 8

saya minta satu mangkuk baru
abang dapat banyak baju lama
kita ambil banyak kotak besar
dia ambil semua kotak lama
kita beri enam meja hijau
abang nampak empat kunci hitam
ayah minta lapan topi putih
adik minta satu topi hijau
kakak dapat tujuh baju mahal
ibu ada enam buku hijau
nenek mahu satu kunci baru
ibu minta lapan topi cantik
ibu dapat enam baju merah
ayah minta satu topi putih

List 15

kita ambil lima kotak putih
dia nampak empat mangkuk lama
ibu perlu semua lampu kecil
kami perlu lapan lampu kecil
ayah ambil lapan kotak cantik
saya nampak enam mangkuk hijau
kami ada empat buku kecil
ibu perlu empat lampu merah
abang ambil satu kotak baru
abang ambil semua kotak kecil
kita beri lima meja besar
ibu dapat lapan baju merah
abang ambil dua kotak hitam
dia suka dua pisau besar
nenek beri tujuh meja cantik
kakak ada enam buku hijau
nenek ambil lapan meja cantik
abang suka dua pisau besar
adik ada lapan buku cantik
saya nampak tiga mangkuk baru
ibu dapat lapan baju cantik
ayah perlu enam lampu putih
dia beri semua meja lama
ibu suka tiga lampu merah
nenek perlu empat lampu cantik
ayah minta lima topi putih
nenek suka empat pisau cantik
kakak ada satu buku mahal
ibu dapat tiga lampu merah
ayah nampak enam mangkuk putih

nenek mahu semua kunci kecil
kakak bagi lapan bola mahal
nenek mahu banyak kunci cantik
kakak dapat tiga baju mahal
saya minta tiga topi merah
ayah suka tiga pisau putih
kita suka semua pisau besar
kita nampak enam mangkuk hijau
kakak beri tujuh kunci mahal
kami ada semua buku kecil
adik dapat banyak baju lama
kita ada empat buku besar
adik suka enam mangkuk hijau
kami suka semua mangkuk kecil
saya suka semua pisau baru
saya dapat satu baju baru
dia beri dua meja besar
abang nampak banyak mangkuk hitam
nenek mahu lapan meja cantik
ibu suka empat pisau merah
abang minta tujuh topi hitam

APPENDIX E: Test lists for MMST-AV in BN

List 1
nenek nampak tiga mangkuk cantik
ayah bagi tujuh bola mahal
ayah ambil lima kotak putih
ibu ada tujuh buku merah
dia suka banyak kotak lama
abang dapat satu baju baru
kami ada tiga buku kecil
abang minta tiga topi hitam
kami suka banyak pisau kecil
kakak ambil lima kotak putih
ibu nampak lima mangkuk putih
kita suka empat pisau hitam
kita suka satu pisau besar
abang mahu empat kunci hitam
ibu suka semua pisau merah
ibu ambil tujuh kotak mahal
adik ambil lapan kotak hijau
kakak perlu enam lampu hijau
saya beri banyak meja lama
nenek mahu lapan topi cantik
abang suka tujuh pisau hitam
ayah minta tujuh topi mahal
ibu nampak tiga kunci merah
ibu ambil semua kotak kecil
nenek mahu tujuh kunci cantik
saya ada empat buku baru
ibu ambil tiga kotak merah
ayah dapat dua baju besar
dia perlu dua lampu besar
kita dapat empat baju besar

List 2
ibu suka lima pisau putih
ibu nampak semua mangkuk kecil
ibu ambil semua kotak merah
kita suka empat pisau besar
ayah suka empat pisau putih
ayah mahu lapan kunci putih
kakak beri tujuh baju mahal
ayah dapat empat baju putih
ayah beri tiga meja merah
kita mahu dua pisau besar
abang dapat satu baju hitam

List 9
nenek nampak satu mangkuk cantik
kita minta satu topi besar
ayah beri enam meja hijau
kakak minta lapan topi mahal
kita perlu dua lampu besar
nenek mahu empat kunci cantik
ibu nampak tiga topi merah
ayah mahu tujuh kunci putih
kami ada banyak buku lama
saya ambil enam kotak hijau
kami beri banyak meja kecil
ayah nampak lapan mangkuk putih
dia ada lima buku lama
abang dapat lapan baju hitam
kakak minta tujuh pisau mahal
abang ada enam buku hijau
adik bagi tujuh bola mahal
saya beri tujuh meja mahal
ibu nampak tujuh mangkuk mahal
ibu ada satu buku merah
saya nampak satu topi baru
adik ambil semua kotak hijau
dia ada tiga buku lama
abang ambil tujuh kotak hitam
dia beri satu meja baru
dia minta dua topi lama
saya suka banyak pisau baru
kami bagi dua bola kecil
abang ambil tujuh kotak mahal
ibu ada dua buku besar

List 10
adik ada tujuh buku hijau
kami suka satu pisau baru
kakak dapat banyak baju lama
ibu minta tujuh topi mahal
adik dapat dua baju hijau
kami nampak empat mangkuk kecil
abang perlu tujuh lampu hitam
nenek nampak banyak mangkuk lama
ayah ambil banyak kotak lama
saya ambil tujuh kotak baru
ayah beri tujuh meja mahal

kita mahu semua kunci kecil
kami ambil enam kotak hijau
ayah ambil tujuh kotak mahal
ibu nampak empat mangkuk merah
kita ada semua buku kecil
kakak beri empat meja mahal
ayah ada tiga buku putih
kita minta lapan topi cantik
abang ada lima buku putih
kami perlu tiga lampu merah
kakak nampak tujuh kotak mahal
kita nampak lima mangkuk putih
kami ada dua buku besar
ibu ada banyak buku merah
saya nampak tiga mangkuk merah
abang mahu tujuh kunci mahal
nenek minta tujuh topi mahal
kita ambil semua kotak kecil
ayah perlu lapan lampu cantik

List 3

ibu ada tiga buku merah
ayah perlu empat lampu putih
dia suka tujuh pisau mahal
ibu bagi empat bola hitam
ayah perlu lima lampu putih
ibu nampak satu mangkuk baru
ayah ambil dua kotak putih
adik nampak dua mangkuk besar
nenek ambil lapan kunci cantik
kita dapat enam baju hijau
ayah suka lima kunci putih
ayah dapat tiga baju putih
dia nampak semua mangkuk kecil
abang minta tujuh topi mahal
adik nampak enam mangkuk hijau
abang nampak tujuh mangkuk mahal
adik beri tujuh meja mahal
saya minta banyak topi lama
kami ambil tujuh kotak mahal
ayah minta empat topi hitam
ibu minta lapan topi merah
abang suka empat kunci hitam
ibu minta banyak topi lama
ayah suka lapan pisau putih
kita ambil lapan kotak cantik
ayah ada lapan buku cantik

ibu minta lima topi putih
dia beri banyak meja lama
dia nampak tujuh mangkuk mahal
abang perlu empat lampu hitam
abang ada tujuh buku mahal
kakak perlu lima lampu putih
nenek ambil empat kotak hitam
adik suka dua pisau besar
abang mahu lima kunci hitam
dia suka satu pisau baru
dia nampak lapan mangkuk cantik
abang minta satu topi hitam
kakak mahu lima kunci mahal
ayah perlu enam lampu hijau
nenek minta lapan mangkuk cantik
kita nampak lapan mangkuk besar
kakak perlu enam lampu mahal
adik ambil satu kotak hijau
kami nampak tiga mangkuk merah

List 11

kami nampak tujuh mangkuk mahal
kami ada tiga buku merah
kami bagi banyak bola lama
kita ambil satu kotak besar
kami ada tujuh buku kecil
abang dapat enam baju hijau
kami ambil enam kotak kecil
ayah suka satu pisau baru
kita nampak dua pisau besar
nenek suka lapan baju cantik
kakak dapat lapan baju cantik
saya beri enam meja hijau
ibu suka satu pisau merah
saya perlu semua lampu kecil
kami perlu enam lampu kecil
adik suka semua pisau kecil
ibu nampak tiga kotak merah
dia ada empat buku hitam
kakak nampak tiga mangkuk merah
kami suka empat pisau kecil
kakak mahu lima kunci putih
dia minta banyak topi lama
nenek bagi enam bola cantik
kakak nampak lima mangkuk mahal
kakak suka dua pisau mahal
abang mahu empat pisau hitam

ayah suka dua pisau besar
ibu suka tujuh pisau merah
adik mahu enam kunci hijau
nenek minta tiga topi merah

List 4

dia ambil lapan kotak cantik
ayah ada semua buku putih
adik minta tujuh topi hijau
abang nampak empat pisau hitam
kakak ambil banyak kotak lama
nenek ada enam buku hijau
kakak minta empat topi hitam
adik beri lapan meja cantik
ibu ambil empat kotak merah
nenek suka dua pisau cantik
kakak suka semua pisau kecil
abang nampak lapan mangkuk hitam
kakak minta enam topi hijau
dia nampak dua mangkuk besar
kakak suka lima pisau putih
ibu dapat tujuh baju mahal
nenek suka banyak pisau lama
saya suka satu kunci baru
nenek mahu lima kunci cantik
kita mahu dua lampu besar
kami perlu lapan lampu cantik
nenek perlu semua lampu cantik
ayah mahu lima pisau putih
saya nampak satu kunci baru
ayah nampak tujuh mangkuk putih
kami suka dua pisau besar
dia suka banyak mangkuk lama
saya ambil tujuh kotak mahal
kakak minta satu topi mahal
dia ada dua buku lama

List 5

kakak mahu tiga kunci mahal
kakak ambil semua kotak mahal
nenek suka banyak pisau cantik
kami perlu empat lampu kecil
dia dapat tiga baju merah
ibu perlu tujuh lampu mahal
saya dapat banyak baju lama
saya nampak banyak mangkuk baru
adik mahu lima kunci hijau
adik nampak tiga mangkuk hijau

nenek suka lapan kotak cantik
ayah suka semua pisau putih
kami ambil semua mangkuk kecil
kakak bagi empat bola hitam

List 12

kakak nampak semua mangkuk mahal
kami dapat semua baju kecil
kakak dapat enam baju mahal
nenek ambil enam kotak cantik
ayah bagi empat bola hitam
dia suka semua pisau kecil
kakak ambil enam kotak hijau
nenek ambil tujuh kotak mahal
saya perlu lima lampu putih
adik beri empat meja hijau
kami mahu banyak kunci lama
kami ada lapan buku cantik
kakak ada tujuh buku mahal
ayah dapat lapan baju putih
nenek suka satu pisau cantik
ibu minta enam topi merah
adik nampak banyak mangkuk lama
dia ada dua buku besar
kakak nampak tujuh topi mahal
ibu minta lima topi merah
dia ada empat buku lama
nenek ada lima buku cantik
nenek perlu tujuh lampu cantik
nenek suka lapan mangkuk cantik
kakak minta tiga topi merah
dia ambil lima kotak lama
ibu nampak lapan mangkuk merah
saya minta lima topi baru
ibu minta empat topi merah
kita beri dua meja besar

List 13

kami beri banyak meja lama
nenek minta lima topi cantik
ibu suka lapan pisau merah
ibu minta tiga mangkuk merah
ibu perlu tiga lampu merah
abang mahu empat topi hitam
ibu dapat empat baju hitam
adik ambil semua kotak kecil
kami perlu empat lampu hitam
nenek minta satu topi baru

ayah ada lima buku putih
kakak perlu empat lampu mahal
kakak bagi enam bola mahal
abang mahu empat meja hitam
kita nampak semua mangkuk besar
adik beri lima meja hijau
nenek minta enam topi hijau
saya suka satu pisau baru
nenek suka tujuh pisau cantik
nenek ambil enam kotak hijau
ayah nampak tujuh mangkuk mahal
kakak nampak dua mangkuk mahal
nenek ambil banyak kotak cantik
abang suka satu pisau hitam
adik suka empat pisau hitam
nenek dapat empat baju cantik
abang suka banyak pisau lama
adik suka satu pisau baru
ayah ada semua buku kecil
ibu suka dua pisau besar

List 6

ayah minta empat topi putih
saya nampak empat mangkuk baru
ayah mahu tiga kunci putih
dia ada enam buku hijau
ibu ambil lapan kotak merah
kita bagi banyak bola lama
kakak perlu tujuh lampu mahal
adik minta enam kunci hijau
kita nampak dua meja besar
abang dapat tujuh baju mahal
kami nampak banyak mangkuk kecil
kakak nampak empat mangkuk mahal
saya ambil lapan kotak baru
kami ada tujuh buku mahal
saya dapat dua baju besar
kakak dapat lapan baju mahal
nenek beri lapan kotak cantik
kakak nampak enam mangkuk mahal
kami nampak empat mangkuk hitam
kakak dapat tiga baju merah
kita minta banyak topi besar
kami beri empat meja kecil
ayah suka lima pisau putih
kakak suka empat pisau hitam
abang suka empat meja hitam

nenek mahu enam kunci hijau
adik dapat tiga baju hijau
ibu nampak tujuh mangkuk merah
nenek nampak tujuh mangkuk cantik
kakak ambil semua kotak kecil
abang minta tiga topi merah
kita ambil tujuh kotak mahal
abang bagi empat bola hitam
adik suka lima pisau putih
kita minta dua topi besar
kami nampak lima mangkuk putih
kami ambil satu kotak baru
dia mahu satu kunci lama
adik nampak lapan mangkuk hijau
adik dapat lapan baju cantik
abang perlu lima lampu hitam
kita beri satu meja baru
abang minta lapan topi hitam
kakak mahu semua kunci kecil
nenek dapat tujuh baju mahal

List 14

kita bagi semua bola besar
ayah ada satu buku putih
kami perlu lima lampu kecil
kakak ambil tiga kotak mahal
kita nampak empat mangkuk besar
kita suka satu pisau baru
abang mahu semua kunci kecil
dia nampak tiga mangkuk merah
nenek nampak enam mangkuk cantik
abang ada semua buku hitam
dia minta tiga topi lama
kakak nampak enam mangkuk hijau
nenek suka tiga pisau merah
adik mahu lapan kunci hijau
adik dapat dua baju besar
ibu minta empat topi hitam
nenek perlu lima lampu putih
adik beri semua meja hijau
kakak dapat enam baju hijau
abang dapat empat lampu hitam
kakak nampak semua mangkuk kecil
kakak mahu empat kunci hitam
dia nampak enam mangkuk hijau
kakak mahu lapan kunci cantik
adik minta lapan topi hijau

dia beri enam meja lama
abang ada lapan buku hitam
saya nampak dua mangkuk baru
saya nampak satu pisau baru
abang dapat lima baju hitam

List7

ayah perlu tujuh lampu mahal
kami mahu lapan kunci kecil
kami ada lima buku kecil
adik beri empat meja hitam
abang mahu empat lampu hitam
ibu minta dua topi merah
ayah ambil enam kotak putih
kita ambil semua kotak besar
kami suka tujuh pisau kecil
nenek minta lapan topi cantik
nenek nampak lima mangkuk putih
abang suka empat kotak hitam
ayah ada lapan buku putih
ayah dapat lima baju putih
nenek beri lima meja putih
abang suka satu pisau baru
adik minta enam topi hijau
abang dapat dua baju hitam
ibu bagi tujuh bola mahal
ayah suka tujuh pisau putih
kakak ada lapan buku cantik
kakak bagi tujuh bola mahal
abang ambil lapan kotak hitam
dia ada tujuh buku mahal
adik ambil dua kotak besar
kakak nampak tujuh meja mahal
kakak ada lima buku mahal
abang ambil enam kotak hitam
nenek ada tujuh buku cantik
kakak minta lima topi mahal

List 8

nenek nampak lapan topi cantik
kakak perlu lapan lampu cantik
ayah ambil empat kotak putih
kita nampak tiga mangkuk besar
ayah mahu semua kunci kecil
kita nampak dua topi besar
ayah ada empat buku hitam
nenek suka lima pisau cantik
kita ambil satu kotak baru

kita suka banyak pisau lama
saya beri satu kotak baru
kakak minta satu topi baru
ibu suka dua pisau merah
adik ambil dua kotak hijau

List 15

ayah perlu empat lampu hitam
kakak suka banyak pisau mahal
saya minta tujuh topi mahal
kakak beri lima meja mahal
adik minta tujuh topi mahal
kami mahu lapan kunci cantik
kami ambil banyak kotak kecil
kita nampak banyak mangkuk lama
ayah dapat tujuh baju putih
saya nampak empat mangkuk hitam
kakak mahu satu kunci baru
kakak ada semua buku kecil
adik nampak lima mangkuk hijau
abang ambil tiga kotak merah
nenek ada semua buku kecil
dia dapat tujuh baju mahal
ayah beri banyak meja putih
nenek perlu empat lampu hitam
ayah ada tujuh buku mahal
adik nampak tiga mangkuk merah
ayah beri lima meja putih
nenek beri lima meja cantik
ayah ambil lima meja putih
adik nampak empat mangkuk hitam
nenek suka semua pisau kecil
dia beri empat meja hitam
adik ada tiga buku hijau
kakak suka satu pisau baru
ayah mahu tiga kunci merah
adik minta empat topi hitam

adik ambil empat kotak hitam
abang minta lima topi hitam
dia nampak lima mangkuk putih
dia ambil empat kotak hitam
abang nampak dua mangkuk hitam
kami suka empat pisau hitam
saya minta dua topi besar
ibu nampak tiga mangkuk merah
abang nampak semua mangkuk kecil
abang perlu enam lampu hijau
ayah perlu semua lampu kecil
adik ambil satu kotak baru
kakak beri lapan meja cantik
kakak ambil empat kotak hitam
kami beri dua meja kecil
abang suka tiga pisau merah
kita beri semua meja besar
kita bagi satu bola baru
kakak ambil lapan kotak cantik
kami mahu satu kunci baru
adik beri dua meja besar

APPENDIX F: Supporting documents

- Human Ethics Committee approval letter
- Approval letter to conduct research in Malaysia
- Participant information sheet
- Participant consent sheet

UNIVERSITY OF CANTERBURY

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2013/27

8 May 2013

Saiful Jamaluddin
Department of Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Saiful

The Human Ethics Committee advises that your research proposal "Development of validation for matrix-type Malay speech-in-noise tests for hearing screening and diagnostics" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 2 May 2013.

Best wishes for your project.

Yours sincerely

Lindsey MacDonald
Chair
University of Canterbury Human Ethics Committee

UNIT PERANCANG EKONOMI
Economic Planning Unit
JABATAN PERDANA MENTERI
Prime Minister's Department
BLOK B5 & B6
PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN
62502 PUTRAJAYA
MALAYSIA

EPU

Telefon : 603-8872 3333

Ruj. Tuan: Your Ref.:

Ruj. Kami: Our Ref.: UPE: 40/200/19/2996

Tarikh: Date: 6 June 2013

SAIFUL ADLI JAMALUDDIN
46 Jalan Jasa 1
Taman Jasa, Sg. Tua,
68100, Batu Caves
Selangor
Email: suosaiful@gmail.com

APPLICATION TO CONDUCT RESEARCH IN MALAYSIA

With reference to your application, I am pleased to inform you that your application to conduct research in Malaysia has been *approved* by the **Research Promotion and Co-Ordination Committee, Economic Planning Unit, Prime Minister's Department.** The details of the approval are as follows:

Researcher's name :	**SAIFUL ADLI JAMALUDDIN**
Passport No. / I. C No:	**800218-14-5145**
Nationality :	**MALAYSIAN**
Title of Research :	**"DEVELOPMENT AND VALIDATION OF MATRIX TYPE MALAY SPEECH TEST FOR HEARING SCREENING AND DIAGNOSTICS"**

Period of Research Approved: **3 YEARS**

2. Please collect your Research Pass in person from the Economic Planning Unit, Prime Minister's Department, Parcel B, Level 4 Block B5, Federal Government Administrative Centre, 62502 Putrajaya and bring along two (2) passport size photographs. You are also required to comply with the rules and regulations stipulated from time to time by the agencies with which you have dealings in the conduct of your research.

Information Sheet

Full Project Title	:	*Development and validation of Malay speech-in-noise tests*
Principal Researcher	:	Saiful Adli Jamaluddin, Ph.D student Department of Communication Disorders
Research Supervisor	:	Dr. Greg O'Beirne, Senior Lecturer Department of Communication Disorders

This study is part of a project to produce hearing tests in the Malay language. Part of the project aims to make hearing screening more available to individuals subject to financial, geographical, social or other factors that limit their access to a hearing professional. This internet or telephone-based test will provide an indication as to whether a person should seek professional opinion regarding their hearing, or if their hearing is normal. The second part of the project aims to create a speech-in-noise test to supplement the information gathered from more traditional audiometric tests.

I am carrying out this research as part of my PhD. Testing will be carried out at the University of Canterbury (either in the Audiology clinics of the Department of Communication Disorders, or the Audiology laboratory in Rutherford 801) or at the Hearing & Speech Clinic, International Islamic University Malaysia. Each session will take 45 minutes to one hour.

Your ears will first be examined, and you will be asked for your history of the health of your ear and hearing. You will then undergo a hearing check (if you have not provided an audiologist-completed audiogram dated within six months), and will be informed of the results of this. I am more than happy to write a letter summarising the results if you would like to follow up on this with your GP or an audiologist.

In part one of the test, you will listen through headphones or a telephone handset to a series of digit "triplets" (e.g. 6-7-9) spoken in Malay, in the presence of background noise. The digits will vary in loudness, and may be difficult to hear at times. After listening to a triplet, please key in the digits you have heard.

In part two, you will watch video of short sentences being read in noise. Sometimes you will only see or hear the sentences being read. In each situation, you are to choose the sentence you heard by pressing words on a screen.

This study will enable me to produce a new way of testing speech perception for the hearing impaired population in Malaysia .The information I obtain from you will be used as a reference to develop the materials further so that it can be used as a hearing screening and diagnostic tool. You may receive a copy of the project results by contacting the researcher at the conclusion of the project.

If you have any queries, I am happy to answer these. You are also welcome to contact me by phone or email should you have questions at a later date.

I have provided a consent form for you to sign prior to participating in this study.

Signing this indicates your understanding that the data collected in this study will not be anonymous, but it will be confidential, and only viewed by people directly involved in this study (those listed below). Participation is voluntary and you have the right to withdraw at any stage without penalty. If you withdraw, I will remove all of the information relating to you.

The project has been reviewed and approved by the University of Canterbury Human Ethics Committee
(HEC 2013/27).

For your own reference, please take this form away with you.

With thanks,

Saiful Adli Jamaluddin
PhD Student
Department of Communication Disorders
University of Canterbury
Email: saiful.jamaluddin@pg.canterbury.ac.nz
Phone: +642 219 13199

Greg O'Beirne, PhD
Research supervisor & Senior Lecturer in Audiology
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: gregory.obeirne@canterbury.ac.nz
Phone: +64 3 364 2987 ext. 7085

Alternatively, if you have any complaints, please contact the Chair of the University of Canterbury Human ethics committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz), phone: +64 3 364 2987.

Consent Form

<u>Consent Form for Persons Participating in Research Studies</u>

Tick (✓) the appropriate site of study

1. University of Canterbury, Communication disorder research facility ☐
2. Hearing & Speech Clinic, International Islamic University Malaysia ☐

Full Project Title: *Development and validation of Malay speech-in- noise tests.*

I have read, or have had read to me in my first language, and I understand the Participant Information.

I, ____ agree to participate in this project according to the conditions in the Participant Information.

I will be given a copy of Participant Information and Consent Form to keep.

The researcher has agreed not to reveal the participant's identity and personal details if information about this project is published or presented in any public form.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.

I understand the risks associated with taking part and how they will be managed.

I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.

I understand that I can contact the researcher or supervisor for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

By signing below, I agree to participate in this research project.

Signature	Date
……………………….	…………………..

Note: All parties signing the Consent Form must date their own signature. Please return the consent form to the researcher before you actively participate in this research.